

# Um Novo Estimador Exponencial Por Partes da Curva de Sobrevivência: Um Estudo Comparativo

Fabíola Eugênio Arrabaça Moraes

Orientador: Prof. Dr. Francisco Louzada Neto

Co-orientador: Prof. Dr. Benedito Galvão Benze

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos

2006

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

M827ne

Moraes, Fabíola Eugênio Arrabaça.

Um novo estimador exponencial por partes da curva de sobrevivência: um estudo comparativo / Fabíola Eugênio Arrabaça Moraes. -- São Carlos : UFSCar, 2006.  
145 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2006.

1. Estatística matemática. 2. Análise de sobrevivência (Biometria). 3. Confiabilidade. 4. Bootstrap (Estatística). 5. Função de risco. I. Título.

CDD: 519.54 (20<sup>a</sup>)

## *Dedicatória*

Dedico este trabalho ao meu esposo José Henrique e aos meus filhos Neto e Vanessa pelo amor, compreensão e incentivo ao longo de toda esta etapa. Amo vocês!!!

"Crescer é sempre difícil. Implica em perder coisas para ganhar outras, em ter coragem para admitir os erros, trocar certezas por dúvidas, questionar o impossível, mas realizar o possível."

Clarice Lispector

## *Agradecimentos*

*Agradeço,*

*à Deus que sempre esteve presente em todas as etapas da minha vida e que me presenteou com uma família maravilhosa, que me ensinou por gestos e palavras, o significado das palavras: confiança, respeito, humildade, ajuda, apoio, proteção, compreensão, incentivo, amor, responsabilidade e sinceridade,*

*ao meu esposo, José Henrique Moraes, aos meus filhos, José Moraes Neto e Vanessa Moraes, a minha mãe, Maria de Lourdes, aos meus avós, Aparecida Antônia (in memoriam) e Felice Stropa, as minhas irmãs, Fabiana, Flaviane e Fernanda e a todos os meus familiares pelo amor, pela paciência, pela compreensão, pelo afeto, pelo incentivo, sempre presentes ao longo desta jornada...*

*ao professor Dr. Francisco Louzada Neto por ter acreditado em mim, de quem aprendi, muito mais que o conhecimento necessário para a realização desse trabalho, aprendi com sua experiência, sua dedicação, sua amizade, respeito e principalmente sua sabedoria.*

*ao professor Dr. Benedito Galvão Benze por ter acreditado em mim, pelos ensinamentos no desenvolvimento deste trabalho, pelas idéias, pela dedicação, pela paciência, compreensão, amizade, respeito e principalmente pelo exemplo de dedicação e disciplina ao trabalho.*

*a todos os professores, por terem sido mediadores dos meus conhecimentos, aos funcionários pelos momentos de convívio,*

*ao professor Dr. Lael Almeida de Oliveira e a professora Dra. Tereza Cristina Dias pelas correções em meu exame de qualificação,*

*a todos os colegas de pós-graduação e aos amigos da minha turma do mestrado, pelos momentos de amizade, companherismo, pela troca de experiências e, sobretudo, pela possibilidade de me proporcionar momentos de grande felicidade,*

*à Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), pelo apoio financeiro,*

*a todos os amigos que, de alguma forma, contribuíram para a realização deste trabalho.*

*Obrigada a todos!*

*"A ilusão dos grandes avanços pode nos confundir, se esquecermos dos passos que os precederam"*



# *Resumo*

Neste trabalho, revisamos os principais conceitos básicos de análise de sobrevivência e confiabilidade, sob a abordagem da inferência clássica. Relacionamos os mais conhecidos estimadores não paramétricos das funções de sobrevivência, com dados censurados. Suas estimativas são calculadas e comparadas em um exemplo com dados reais de transplantes renais humanos. Apontamos alguns problemas que, muitas vezes, deixaram de ser considerados na estruturação destes estimadores. Como consequência destes problemas, citamos possíveis distorções nas estimativas, no sentido de sub ou super estimar as principais características de interesse, e levantamos as possibilidades de se estar: expondo a riscos desnecessários pacientes, acompanhados em estudos clínicos, ou submetendo-se empresas a prejuízos financeiros, em estudos de fidelidade de clientes.

Além disso, propomos um novo estimador modificado do tipo exponencial por partes, pela mudança que realizamos na taxa de risco acumulada de Kitchin nos subintervalos formados pelos tempos consecutivos das ocorrências dos eventos de interesse. Finalmente, dentro do conhecimento atualmente disponível, apontamos alternativas na busca de correções para alguns destes problemas e realçamos a necessidade da elaboração de estimadores menos vulneráveis a eles.

**Palavras-chave:** Análise de Sobrevivência, Dados Censurados, Funções de Sobrevivência, Funções de Risco, Inferência Clássica, Estimadores Não Paramétricos, Riscos Competitivos.

# *Abstract*

In this work we revise the main basic concepts on survival analysis and reliability. We present the most known nonparametric estimators of the survival function in the presence of censored data. Their estimates are calculated and compared in an real data set on human renal transplants. We point some problems that are not usually considered in the structure of these estimators. As consequence of such problems we can observe possible distortions in the estimates the direction of sub or over estimating the main characteristics of interest. Some alternatives are pointed out for finding less vulnerable estimators. Moreover, we propose a new modified piece wise exponential estimator, pointing out its properties.

**Word-key:** Survival Analysis, Censored Data, Survival Functions, Nonparametric Estimators.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Conceitos Gerais de Análise de Sobrevivência</b>	<b>6</b>
2.0.1	Descrição do tempo de sobrevivência . . . . .	6
2.0.2	Censuras . . . . .	6
2.0.3	Função densidade de probabilidade e função de distribuição acumulada . . . . .	8
2.0.4	Função de sobrevivência . . . . .	8
2.0.5	Função de risco . . . . .	9
2.0.6	Função de subsobrevivência . . . . .	10
2.0.7	Os limites das funções de sobrevivência em termos das funções de subsobrevivência . . . . .	11
2.0.8	Função conjunta de sobrevivência . . . . .	11
<b>3</b>	<b>Alguns Procedimentos Não Paramétricos, Segundo a Inferência Clássica, para a Estimação da Curva de Sobrevivência</b>	<b>13</b>
3.1	Notação . . . . .	13
3.2	Um exemplo com dados reais . . . . .	15
3.3	Algumas considerações para o estabelecimento deste tempo de seguimento	16
3.4	O estimador empírico da função de sobrevivência . . . . .	17
3.5	Os estimadores empíricos das funções de sub-sobrevivência . . . . .	18
3.6	O estimador atuarial adaptado para pequenas amostras . . . . .	20
3.7	O estimador de Kaplan-Meier . . . . .	22
3.8	O estimador de Nelson . . . . .	25

---

3.9	O estimador exponencial por partes de Kitchin . . . . .	26
3.10	Comparação das estimativas da curva de sobrevivência, segundo os procedimentos não paramétricos apresentados . . . . .	29
<b>4</b>	<b>Alguns Problemas de Interesse</b>	<b>34</b>
4.1	Outro exemplo com dados reais . . . . .	44
<b>5</b>	<b>Proposta de Um Novo Estimador Modificado do Tipo Exponencial Por Partes de Kitchin</b>	<b>50</b>
<b>6</b>	<b>Procedimentos de Reamostragem para Verificar a Precisão dos Estimadores</b>	<b>62</b>
6.1	O estimador bootstrap . . . . .	63
6.1.1	O estimador bootstrap do erro padrão . . . . .	64
6.2	Estimação intervalar . . . . .	67
6.2.1	Estimação por intervalo via bootstrap . . . . .	67
6.2.2	Intervalos de confiança . . . . .	67
6.3	Procedimento de geração das amostras . . . . .	69
6.4	Resultados do estudo de simulação . . . . .	71
<b>7</b>	<b>Conclusões e Perspectivas Futuras</b>	<b>75</b>
	<b>Referências Bibliográficas</b>	<b>76</b>
<b>A</b>	<b>Simulação</b>	<b>77</b>
<b>B</b>	<b>Simulação</b>	<b>105</b>

# Capítulo 1

## Introdução

A análise de sobrevivência e de confiabilidade consiste dos procedimentos apropriados para o estudo estatístico dos tempos de duração, durante o seguimento de uma amostra de unidades experimentais medidas, desde a entrada de cada uma dessas unidades no estudo, até a sua saída. Essa amostra pode ser constituída, por exemplo, de: pacientes com uma determinada doença, submetidos a um tratamento; clientes utilizando um tipo de cartão de crédito ou componentes industriais submetidos a estudos de durabilidade de equipamentos e sistemas, entre outros. A saída de seguimento pode se dar pela ocorrência: de um evento pré-fixado ou de um evento diferente que acontece antes daquele. No primeiro caso, a saída é referida como uma "falha" e no segundo como uma "censura". Assim, os dados basicamente consistem dos tempos de duração até a ocorrência do evento de interesse (tempos de falha) ou até a saída de seguimento por outra causa (tempos de censura ou dados incompletos).

De um modo geral, procuramos obter um modelo que explique satisfatoriamente os principais aspectos dos dados, permitindo a tomada de decisões rápidas pelos pesquisadores, em relação ao problema estudado. Isso tem estimulado o aprimoramento dos métodos de estimação clássicos ou Bayesianos, paramétricos ou não e na presença ou não dos tempos de censura, da forma da distribuição dos tempos de falha e de outras suas características como média, variância e percentis.

Dentre os trabalhos desenvolvidos com este objetivo são destacados aqui os já bem conhecidos de Kaplan e Meier (1958), em ensaios clínicos; de Nelson (1970), em confiabilidade de sistemas e o de Fleming (1978), que seguem a abordagem clássica;

o de Suzarla e Van-Ryzin (1976), e o de Gamerman (1994) que seguem a abordagem Bayesiana. O estimador de Kaplan e Meier da curva de sobrevivência, referido como estimador produto limite (PL) e desenvolvido a partir do chamado estimador atuarial, Littell (1952), citado em Kaplan e Meier (1958), tem sido muito utilizado em estudos clínicos, enquanto o estimador de Nelson da função risco de falha, tem sido mais empregado em estudos de confiabilidade industrial. No trabalho de Fleming é apresentado um estimador mais generalizado que os de Kaplan-Meier e de Nelson, permitindo as estimativas das probabilidades do tempo de duração ou de sobrevida, na presença de múltiplas causas atuando simultaneamente na população. Mais recentemente, entre outros, surgiram os estudos de Kitchin (1980); Kim e Proschan (1991); Salinas (1992) e Benze (1997). O estimador de Kitchin, conhecido como estimador PEXE (piecewise exponential estimator), tem a mesma forma funcional do estimador de Nelson, mas foi construído com o principal propósito de produzir estimativas alisadas da curva de sobrevivência, em substituição às curvas em forma de escada dos estimadores de Kaplan-Meier e de Nelson. Em seu trabalho de doutorado, Benze, mantendo a mesma forma funcional do estimador de Kitchin e simplificando o cálculo das taxas de falhas, obteve um estimador modificado de Kitchin. Com o seu estimador Bayesiano, Salinas apresentou um método de estimação mais geral que o de Suzarla e Van-Ryzin mas, as suas estimativas também não são curvas alisadas. Ainda, do ponto de vista Bayesiano, Gamerman (1994), propôs um alisamento da função estimada de risco de falhas, do tipo exponencial por partes, elaborando um procedimento que incorporou no estimador a possível dependência dos valores das taxas de risco de um subintervalo para os próximos. Entretanto, no trabalho de Gamerman, o estabelecimento do mecanismo controlador de passagem de informação de um intervalo para o próximo não é tão simples para facilitar futuras aplicações do método em outras aplicações.

Apesar disso, os estimadores implementados nos softwares estatísticos, e mais comumente em uso, continuam a ser: o Atuarial (em estudos de grande escalas), Littell (1952), citado em Kaplan e Meier (1958); o Produto Limite de Kaplan-Meier (1958) e os de Nelson (1970). Mais recentemente, estão sendo usados os estimadores do tipo Exponencial por Partes (como, por exemplo, os de: Kitchin (1980) e o de Kim-Proschan (1991).

Em todos os estimadores descritos acima, existem problemas, decorrentes de

condições simplificadoras nas suas elaborações, que podem estar artificializando os resultados e, em conseqüência, sub ou super estimar as características de interesse. Como conseqüência, podemos expor a riscos desnecessários pacientes submetidos a estudos clínicos ou submetendo-se empresas a prejuízos financeiros, em estudos de fidelidade de clientes ou, ao mesmo tempo, arriscando tanto a segurança das pessoas como as finanças das empresas, em estudos de confiabilidade de componentes ou sistemas.

Dentre tais problemas, podemos citar: i) a suposição irrealística de independência entre os riscos de falha que atuam simultaneamente na amostra; ii) a indefinição no final da curva de sobrevivência estimada, quando a última unidade a abandonar o estudo é uma censura; iii) a não consideração da posição dos tempos censurantes dentro dos subintervalos entre dois pontos consecutivos de falha, no processo de estimação neste subintervalo; iv) as estimativas dadas por curvas decrescentes em forma de escada, quando se espera que o modelo teórico seja uma curva contínua decrescendo até zero; v) a vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier, a pequenas perturbações provocadas no conjunto de dados, por exemplo: quando um médico registra inadvertidamente, para um indivíduo da sua amostra, um tempo maior para a ocorrência de sua morte, do que o valor que realmente aconteceu, que denominamos de perturbação do Tipo I, ou quando, erroneamente, é registrado um tempo de sobrevida como censura quando na realidade ocorreu uma falha, que denominamos de perturbação do Tipo II); vi) a inviabilidade nos estimadores atuais em transferir, para a taxa de risco em um dado subintervalo, as informações contidas nos subintervalos anteriores; vii) a não possibilidade, em alguns destes estimadores, da estimação direta em pontos de empates (isto é, quando dois ou mais indivíduos falham ou são censurados ou falham e são censurados em um mesmo tempo); viii) os estimadores consideram que todos os indivíduos entram em um mesmo instante no estudo, o que geralmente não é verdade. Os indivíduos são admitidos em tempos diferentes e isso não é levado em conta no processo de elaboração dos estimadores; ix) qual estimador apresenta melhor desempenho dentre os mais utilizados; além de outras complicações a serem ainda detectadas.

Em um primeiro momento, podemos procurar identificar, dentre os estimadores já consolidados, e para determinadas situações específicas (como, por exemplo: em ensaios clínicos) aquele(s) que seja(m) menos vulnerável(eis) aos problemas já levantados.

O objetivo deste trabalho é estudar especificamente os problemas ix) qual estimador apresenta melhor desempenho dentre os mais utilizados; v) a vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier, a pequenas perturbações provocadas no conjunto de dados, por exemplo: quando um médico registra inadvertidamente, para um indivíduo da sua amostra, um tempo maior para a ocorrência de sua morte, do que o valor que realmente aconteceu, que denominamos de perturbação do Tipo I, ou quando, erroneamente, é registrado um tempo de sobrevivência como censura quando na realidade ocorreu uma falha, que denominamos de perturbação do Tipo II e propor um novo estimador não paramétrico para a curva de sobrevivência, sob dados censurados.

Devemos ressaltar que um dos objetivos deste trabalho foi estudar o problema da vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier, e verificada por nós para os demais estimadores estudados, a pequenas perturbações provocadas no conjunto de dados. Neste contexto, observamos que para a perturbação Tipo I, por exemplo: quando um médico registra inadvertidamente, para um indivíduo da sua amostra, um tempo maior para a ocorrência de sua morte, do que o valor que realmente aconteceu, o estimador de Kitchin foi o mais resistente. Desta forma apresentamos uma solução para problema que Oakes (1993) cita em seu artigo.

Para a perturbação Tipo II, por exemplo: quando, erroneamente, é registrado um tempo de sobrevivência como censura quando na realidade ocorreu uma falha, o estimador proposto  $mk$  apresentou-se melhor depois da metade da distribuição, nas as configurações estudadas. Sob esta perturbação o estimador  $km$  foi o menos resistente.

Concluimos que a introdução das perturbações não alteraram significativamente o bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

No Capítulo 2 definimos os principais conceitos da análise de sobrevivência. Descrevemos também algumas relações entre as funções que descrevem o comportamento do tempo de sobrevivência. No Capítulo 3, usando uma notação comum, apresentamos alguns procedimentos não paramétricos, segundo a inferência clássica, para a estimação da curva de sobrevivência. Em um exemplo com dados clínicos reais de transplantes renais humanos, ilustramos estes procedimentos de estimação não paramétricos.

No Capítulo 4 apresentamos um estimador modificado do tipo exponencial por

partes de Kitchin que estamos propondo neste trabalho, denominado de estimador modificado de Kitchin e comparamos suas estimativas com os demais estimadores apresentados no Capítulo 3.

No Capítulo 5 observamos o efeito das perturbações Tipo I e II, em dois conjuntos de dados reais, de transplantes renais e transplantes de medula óssea, com o objetivo de comparar o comportamento das estimativas segundo os estimadores estudados. No Capítulo 6, comparamos, mediante os estudos de simulação bootstrap, o comportamento geral destes estimadores na situação normal (de não perturbação) e com essas perturbações.

No Capítulo 7 apresentamos as conclusões e perspectivas futuras.

# Capítulo 2

## Conceitos Gerais de Análise de Sobrevivência

Neste capítulo, apresentamos algumas definições básicas da análise de sobrevivência e confiabilidade de sistemas. Descrevemos também algumas relações entre as funções que descrevem o comportamento do tempo de sobrevivência.

### 2.0.1 Descrição do tempo de sobrevivência

O comportamento da variável aleatória contínua tempo de sobrevivência relativamente a um evento específico  $e$ ,  $T_e \geq 0$ , pode ser expresso por quatro funções estatisticamente equivalentes, tais que se uma delas for especificada, as outras podem ser encontradas. Entre elas temos, a função densidade de probabilidade  $f_e(t)$  e/ou a função de distribuição acumulada  $F_e(t)$  e/ou a função de sobrevivência  $S_e(t)$  e/ou a função de risco  $\lambda_e(t)$ . Estas funções são utilizadas na prática para descrever os diferentes aspectos da distribuição do tempo de sobrevivência para conjuntos específicos de dados.

### 2.0.2 Censuras

Um fator complicador que frequentemente ocorre em estudos da análise de sobrevivência e confiabilidade é a possibilidade da saída das unidades experimentais do experimento por causas diferentes daquela escolhida para estudo. É o que denominamos censuras, gerando os chamados dados censurados. Nestes casos, dispomos somente dos

registros de seguimentos até um certo instante que correspondem, por exemplo, as datas dos últimos retornos de pacientes, com o qual perdemos o contato, ou ao momento em que os pacientes foram retirados do seguimento, logo após o pesquisador tomar conhecimento dos seus envolvimento com a ocorrência de um evento estranho à pesquisa. Da mesma maneira, este tipo de problema pode estar presente em dados industriais, quando somente uma proporção de componentes em estudo falham, por causa de um determinado tipo de defeito, até o final do experimento.

Em geral, a censura ocorre à direita isto é, para as unidades censuradas, sabemos que elas duraram pelo menos até um certo tempo  $t$  e saíram do seguimento por causas diferentes daquela de interesse. Mas podem ocorrer também censuras à esquerda, por exemplo, não se sabe exatamente a partir de qual tempo os indivíduos adquiriram uma determinada doença.

Os processos de censuras são comumente classificados em três tipos: tipo I, tipo II e *aleatório*.

A censura do tipo I ocorre em experimentos que têm período de duração fixado. Por exemplo, suponhamos que os  $n$  componentes colocados em teste só sejam acompanhados durante um certo período de tempo, após o qual o experimento é encerrado. Aquelas unidades que permaneciam em funcionamento no instante do término do experimento não terão mais os seus tempos de falhas observados; no entanto, é sabido que não falharam até o final do ensaio. Portanto, os tempos de vida das unidades experimentais só serão observados se forem menores ou iguais ao tempo de execução do experimento; caso contrário, consistem de observações censuradas.

A censura do tipo II ocorre em experimentos que terminam a partir do momento em que um certo número de falhas é atingido. No exemplo anterior, suponhamos que o custo dos componentes em teste seja considerável, de modo que o pesquisador só se disponha a perder  $r$  componentes, até o encerramento do ensaio. Nesta situação, se o experimento tem início com  $n$  unidades experimentais, após a  $r$ -ésima falha ( $r \leq n$ ) ocorrerão  $(n - r)$  observações censuradas.

Neste trabalho as censuras que não satisfizerem as definições de censura do tipo I e do tipo II, serão consideradas censuras aleatórias.

As censuras do tipo I e aleatória são mais comuns na área médica, ao passo que a

censura de tipo II é frequentemente mencionada em experimentos na área de engenharia.

### 2.0.3 Função densidade de probabilidade e função de distribuição acumulada

A função densidade de probabilidade é definida como o seguinte limite da densidade de probabilidade, da unidade em estudo vir a falhar devido a ocorrência do evento  $e$  em um subintervalo de tempo  $[t, t + \Delta t)$ , de comprimento  $\Delta t$ :

$$fe(t) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq Te < t + \Delta t)}{\Delta t} \quad (2.1)$$

O termo densidade de probabilidade é usado porque  $P(t \leq Te < t + \Delta t)/\Delta t$  é uma densidade. A partir da  $fe(t)$  definimos a função de distribuição acumulada de  $Te$  por,

$$Fe(t) = P[Te \leq t] = \int_0^t fe(u)du, \quad (2.2)$$

onde  $t \in R$ .

### 2.0.4 Função de sobrevivência

A função de sobrevivência ou de confiabilidade  $Se(t)$  é definida como a probabilidade da unidade em estudo somente falhar pela ocorrência do evento e após um certo tempo  $t$ , e é indicada por,

$$Se(t) = P(Te > t) = 1 - Fe(t), \quad (2.3)$$

de modo que  $Se(t) = 1$  quando  $t = 0$  e  $Se(t) = 0$  quando  $t \rightarrow \infty$ .

Podemos mostrar que,

$$fe(t) = -\frac{dSe(t)}{dt} \quad (2.4)$$

### 2.0.5 Função de risco

A função de risco  $\lambda e(t)$  é definida como o seguinte limite da densidade de probabilidade condicional, da unidade em estudo falhar no intervalo de tempo  $[t, t + \Delta t)$ , dado que a mesma não tenha falhado até o tempo  $t$ :

$$\begin{aligned}\lambda e(t) &= \lim_{\Delta t \rightarrow 0} \left[ \frac{P[t < Te \leq t + \Delta t \mid Te > t]}{\Delta t} \right] \\ &= \frac{fe(t)}{Se(t)} \\ &= -\frac{d}{dt} \ln Se(t).\end{aligned}\tag{2.5}$$

Podemos mostrar que a *função risco*  $\lambda e(t)$  pode ser escrita como a razão das funções densidade de probabilidade  $fe(t)$  e de sobrevivência  $Se(t)$  e de (2.4) vemos que a função risco pode ser expressa como a derivada negativa do logaritmo natural da função de sobrevivência.

Ao contrário da função de sobrevivência  $Se(t)$  que se refere à não falha, a função de risco  $\lambda e(t)$  focaliza-se na falha. Uma interpretação da função de risco é a seguinte: " $\lambda e(t)$  é o potencial instantâneo de falha, por unidade de tempo, dado que o indivíduo tenha sobrevivido ao tempo  $t$ ", Kleinbaum (1989).

Devido à probabilidade condicional na sua definição, a função  $\lambda e(t)$  é às vezes chamada "taxa de falha condicional". Notamos que:  $\lambda e(t)$  não é uma probabilidade, pois é definida por uma razão que tem no numerador uma probabilidade e no denominador o comprimento  $\Delta t$  do intervalo de tempo. Essa razão é uma probabilidade por unidade de tempo (isto é, uma taxa). A sua variação não é de 0 e 1 (como acontece com probabilidade), mas de 0 a  $\infty$  e depende de como o tempo é medido.

Uma outra função importante que pode ser obtida de (2.5) é dada pela função de risco acumulada definida por,

$$\begin{aligned}\Lambda e(t) &= \int_0^t \lambda e(u) du \\ &= -\ln Se(t)\end{aligned}\tag{2.6}$$

### 2.0.6 Função de subsobrevivência

A variável de maior interesse em nossos estudos é o tempo potencial de sobrevivência da unidade em estudo quando somente um risco específico de falha  $Re$  fixado estiver agindo contra seu funcionamento. No contexto de riscos competitivos, o que realmente observamos são  $n$  tempos  $T_i$  de duração do seguimento, correspondentes a cada indivíduo  $i$  da amostra estudada, que permaneceu exposto a todos os  $k$  riscos previstos de falha, até a sua saída de observação pela ocorrência de uma das causas  $Ck$ . À saída de interesse é atribuída à ocorrência somente de uma única causa  $Ce$ , que é associada ao único risco  $Re$ , do conjunto de todos os  $K$  riscos considerados.

Dessa forma, o tempo observado  $T_i$ , de duração do  $i$ -ésimo indivíduo da amostra, é o menor dos  $k$ -i-ésimos tempos potenciais de duração  $Tk_i$ , atribuídos a esse indivíduo, entre todos quando se leva em conta isoladamente cada risco a que ele esteja exposto. Essas observações poderão ser incompletas pois a informação efetivamente obtida para cada unidade amostral  $i$  é o par  $(T_i, \delta_i)$ , em que

$$T_i = \min_{1 \leq k_i \leq K} \{Tk_i\}, i = 1, 2, \dots, n \quad (2.7)$$

e  $\delta_i = \delta_{1i}, \delta_{2i}, \dots, \delta_{Ki}$ , onde

$$\delta_{k_i} = \begin{cases} 1, & \text{quando } T_i = Tk_i, i = 1, \dots, N; k_i = 1, \dots, K. \\ 0, & \text{caso contrário} \end{cases} \quad (2.8)$$

Assim, temos:

**Hipótese 1:** "Os tempos de sobrevivência,  $Tk_i$ ,  $k = 1, \dots, K$  são mutuamente independentes".

Fixando um risco  $Re$  de interesse, observamos os  $n$  tempos  $Te_i$  na amostra (os quais, algumas vezes, representamos apenas por  $T$ ) e as correspondentes causas  $Ce_i$  de saídas de segmento (às vezes simplificadas por  $Ce$ ). Nestas condições, podemos reduzir o problema de riscos competitivos ao problema da exposição das unidades amostrais a somente dois riscos. O risco de interesse  $Re$  (com a causa  $Ce$  associada) e o risco de censura  $Rc$ , referindo-se a todas as outras causas  $Cc$ , onde são consideradas as ocorrências de todos os outros riscos. Assim, podemos definir as chamadas *funções de subsobrevivência*

$Se^*(t)$  e  $Sc^*(t)$ .

As funções de subsobrevivência são dadas pela probabilidade de que as durações  $Te_i$  (que simplesmente representamos por  $T$ ) excedam a um tempo  $t$  positivo, pré-estabelecido, e que só sejam finalizadas por uma das causas  $Ck$  aqui ( $Ce$  ou  $Cc$ ).

Assim,

$$Se^*(t) = P\{\min(T) > t; Ce\} \quad (2.9)$$

$$Sc^*(t) = P\{\min(T) > t; Cc\} \quad (2.10)$$

### 2.0.7 Os limites das funções de sobrevivência em termos das funções de subsobrevivência

Na ausência da hipótese de independência entre os riscos  $R_k$  (ou das correspondentes causas  $C_k$ , ou dos correspondentes tempos  $T_k$ ) podemos estabelecer, por Peterson (1976), os seguintes limites para a função de sobrevivência quando  $Re$  é considerado o risco específico de interesse atuando sobre os indivíduos da amostra.

$$Se^*(t) + Sc^*(t) \leq Se(t) \leq Se^*(t) + Fe^*(0) \quad (2.11)$$

onde  $Fe^*(0) = 1 - Se^*(0)$

### 2.0.8 Função conjunta de sobrevivência

A função conjunta de sobrevivência  $S_T(t)$  é definida pela probabilidade de que todas as durações no seguimento excedam ao valor pré-fixado  $t \geq 0$ , sem especificarmos um particular risco de interesse, dentre todos aqueles competindo para a eliminação dos indivíduos, ou seja,

$$\begin{aligned} S_T(t) &= P[T > t] \\ &= P[T_1 > t, T_2 > t, \dots, T_K > t]. \end{aligned} \quad (2.12)$$

Para os dois riscos  $Re$  e  $Rc$ , (2.12) assume a forma bidimensional:

$$S_T(t) = P[Te > t, Tc > t].$$

Maiores detalhes sobre os conceitos gerais da análise de sobrevivência podem ser encontrados em Lawless (1982).

No próximo Capítulo, usando uma notação comum, apresentamos alguns procedimentos não paramétricos, segundo a inferência clássica, para a estimação da curva de sobrevivência. Em um exemplo com dados clínicos reais de transplantes renais humanos, ilustramos estes procedimentos de estimação não paramétricos.

# Capítulo 3

## Alguns Procedimentos Não Paramétricos, Segundo a Inferência Clássica, para a Estimação da Curva de Sobrevivência

Neste capítulo, apresentamos alguns procedimentos não paramétricos de estimação das funções definidas em (2.3), (2.9) e (2.12). Estes procedimentos são calculados e ilustrados em um exemplo com dados clínicos reais.

### 3.1 Notação

A seguir, representamos a sequência ordenada dos tempos observados de ocorrências dos eventos,

$$0 < t_1 < t_2 < t_3 < \dots < t_{j-1} < t_j < \dots < t_{N-1} < t_N, \quad 1 \leq j \leq N \leq n, \quad (3.1)$$

onde o índice " $j$ " é um inteiro indicando a ordem da observação em um seguimento e cada  $t_j$  é o valor observado do tempo  $te$  de falha ou do tempo  $tc$  de censura. O índice  $j$  pode não atingir o valor correspondente ao tamanho  $n$  da amostra devido à possibilidade da ocorrência de empates.

Em seguida, consideramos os subintervalos  $I_m = (te_{m-1}; te_m]$ , determinados por cada dois pontos consecutivos  $te_{m-1}$  e  $te_m$ , de falhas, para  $m = 1, \dots, M$ . A letra minúscula  $m$  especifica a ordem dos valores dos tempos que correspondem somente às falhas devidas a uma causa específica fixada  $Ce$  de interesse. A Figura 3.1 foi construída para esclarecimentos sobre esta notação.

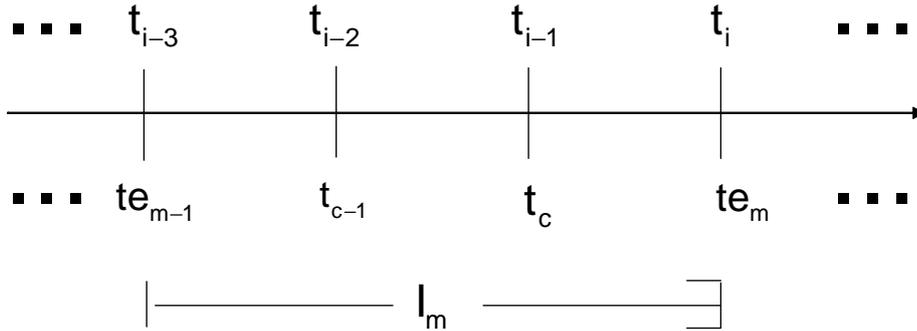


Figura 3.1 Subintervalos  $I_m$  determinados por cada dois pontos consecutivos  $te_{m-1}$  e  $te_m$  de falhas.

Na Figura 3.1:

$t_c = c$ -ésimo tempo de censura (saída devida a outra causa).

$$\Delta t_m = te_m - te_{m-1}$$

A seguir, definimos a notação que será utilizada posteriormante:

$V(te_m)$  = número de unidades em risco imediatamente após o  $(m-1)$ -ésimo tempo observado de falhas.

$V(te_m^-)$  = número de unidades em risco imediatamente antes do  $m$ -ésimo tempo de falhas.

$M(te_m)$  = número de "falhas" registradas no subintervalo  $I_m$  pela causa de interesse, exatamente no tempo  $te_m$ .

$W(I_m)$  = número de censuras registradas no subintervalo  $I_m$ , entre os tempos de falhas.

$\lambda e(t)$  = taxa de risco instantâneo de falha no tempo  $t$  devido à causa de interesse,

$Ce$ , já definida em (2.5).

$\Lambda e(t)$  = taxa de risco acumulado de falha até o ponto  $t$  devido à causa de interesse,  $Ce$ , já definida em (2.6).

A seguir, descrevemos um exemplo com dados reais onde depois ilustraremos os procedimentos de estimação não paramétricos apresentados, a partir da Seção 3.4.

## 3.2 Um exemplo com dados reais

O conjunto de dados que consideramos neste trabalho é de Aquino (1979), os quais se referem a 25 observações, durante o período de 14/02/68 a 31/12/78, de pacientes de ambos os sexos, com idades iguais ou maiores do que 12 anos. Eles apresentaram diagnóstico anátomo-patológico de insuficiência renal crônica terminal, com a realização de tratamento dialítico e posterior alotransplante renal com doador vivo, na Unidade de Transplantes Renais (UTR), do hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, no Campus de Ribeirão Preto.

Após o transplante, foi administrado a cada indivíduo o mesmo esquema terapêutico imunodepressor, com o objetivo de inibir a rejeição, observando-se simultaneamente o tempo de sobrevida do órgão transplantado, isto é, o tempo de funcionamento do "novo" rim, a partir da cirurgia e o motivo que determinou a sua finalização de seguimento. Entre esses foram citados: óbitos diretamente relacionados ao enxerto e o término de acompanhamento ao tratamento (como, por exemplo, as desistências na continuação do tratamento). Em outros estudos, semelhantes a este, existe a possibilidade de também ocorrer: o retorno do paciente à hemodiálise, pela parada de funcionamento do rim transplantado e óbitos não diretamente relacionados ao enxerto.

A seguir, na Tabela 3.1, estão representados a correspondente duração (em dias)

do tempo de seguimento, para cada paciente na amostra estudada.

Tabela 3.1 Representação da duração (em dias) do tempo de seguimento para cada paciente em estudo com (<sup>c</sup>) representando os tempos censurantes.

$t_1=40$	$t_2=43$	$t_3=89$	$t_4=201$	$t_5=215$
$t_6=308$	$t_7=396$	$t_8=512$	$t_9=542$	$t_{10}=956^c$
$t_{11}=970^c$	$t_{12}=1227$	$t_{13}=1272^c$	$t_{14}=1343^c$	$t_{15}=1362^c$
$t_{16}=1424$	$t_{17}=1426^c$	$t_{18}=1434^c$	$t_{19}=1483^c$	$t_{20}=1518^c$
$t_{21}=1570$	$t_{22}=1588^c$	$t_{23}=1609^c$	$t_{24}=1713^c$	$t_{25}=1973$

### 3.3 Algumas considerações para o estabelecimento deste tempo de seguimento

Os primeiros transplantes, nestes pacientes, só foram realizados a partir do início de 1972. Assim consideramos, como período de observação, o intervalo de 14/02/72 a 31/12/78. A duração do tempo do seguimento (em dias), entre as entradas (isto é, o dia da cirurgia), e as saídas de estudo de cada paciente, durante o acompanhamento, pode ser considerada aleatória pelas diferenças na evolução da doença, pela disponibilidade de doador compatível e por vários outros fatores que podem influenciar a resposta ao tratamento. Temos uma amostra constituída de 48% de observações censuradas, devido ao truncamento à direita provocado pela decisão de avaliar o resultado do experimento após o 2512º dia de seguimento dos pacientes. Também, é considerado o esquema de censura tipo I, "não informativo". É importante ressaltar que existe a possibilidade das censuras serem "não informativas", isto é, não se supõe outra informação nos dados censurados da amostra, além da duração do tempo de seguimento e a causa de finalização de observação. Assim, o "mecanismo de produção" de censuras não será explicitamente considerado nos métodos de estimação das probabilidades de sobrevivência.

Além disso, admitimos a suposição de independência entre os tempos potenciais de falhas e censuras, o que pode ser questionado em muitas situações. É o caso, por exemplo, em que o mecanismo de censuras pode conduzir a um prognóstico desfavorável

sobre a sobrevivência dos indivíduos em estudo.

Os dados aqui escolhidos para ilustrar os métodos de estimação da função de sobrevivência são obtidos em estudos nos quais os indivíduos chegam em diferentes tempos durante um período de observação pré-fixado pelo pesquisador, ao invés da pesquisa iniciar-se com um número fixo deles. Como os tempos de entrada de cada paciente, durante o acompanhamento, não são simultâneos, os tempos censurados também podem ser diferentes, ou seja, progressivamente censurados e satisfazem a definição de censura à direita do tipo I (conforme é estabelecido na teoria de análise de sobrevivência e confiabilidade).

O experimento consistiu em observar o tempo de sobrevivência dos pacientes submetidos ao transplante, no decorrer do tempo. A variável de maior interesse aqui foi o tempo potencial de sobrevida do enxerto se somente o risco específico  $Re$  de morte devido à falha do rim transplantado estivesse agindo no indivíduo, a qual indicaremos por  $Te$ . O tempo  $Te$  é a duração que teoricamente ocorreria no acompanhamento de um indivíduo, nas condições em estudo, caso pudéssemos observá-lo sob a influência isolada do risco  $Re$ , até a realização da correspondente causa específica  $Ce$  da sua finalização no seguimento.

### 3.4 O estimador empírico da função de sobrevivência

Uma estimativa da função conjunta de sobrevivência (2.12) é dada pelo estimador empírico:

$$\hat{S}_e(t) = \frac{\text{Número de indivíduos que sobreviveram além de } t}{\text{Número total de indivíduos em estudo}}, \quad (3.2)$$

que é equivalente a expressão

$$\hat{S}_n(t) = \frac{1}{n} \sum_{i=1}^n I(T_i > t), \quad (3.3)$$

em que  $I[\cdot]$  é a função indicadora das durações observadas que excedem um valor especificado  $t \geq 0$ , finalizadas por quaisquer das causas  $Ck$ ,  $k = 1, 2, \dots, K$ .

As estimativas segundo este estimador foram calculadas para o exemplo com dados reais de transplantes renais, já citado na seção 3.2 e a representação gráfica está ilustrada

na Figura 3.2.

### 3.5 Os estimadores empíricos das funções de sub-sobrevivência

Os estimadores das funções de subsobrevivência (2.9) e (2.10) são estabelecidos por

$$\hat{S}_e^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i > t; Ce). \quad (3.4)$$

e

$$\hat{S}_c^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i > t; Cc). \quad (3.5)$$

em que  $\hat{S}_k^*(\cdot)$  denota o estimador empírico e  $I[\cdot]$  é uma função indicadora das durações que excedem o valor especificado  $t \geq 0$  e que são finalizadas somente pela causa  $Ck$  ( $k = e$  ou  $k = c$ )

Notamos que aqui estamos estimando as probabilidades de sobrevivência ao tempo  $t$ , com interesse em um risco específico  $Rk$ , com os  $n$  indivíduos da amostra sujeitos à ação simultânea deste e de todos os outros riscos concorrendo contra a sua permanência em estudo.

A seguir, as estimativas foram calculadas para o exemplo com dados reais de transplantes renais e a representação gráfica está ilustrada na Figura 3.2. Nesta figura também representamos as estimativas da curva de sobrevivência da função conjunta de sobrevivência. Os valores numéricos foram registrados na Tabela 3.2.

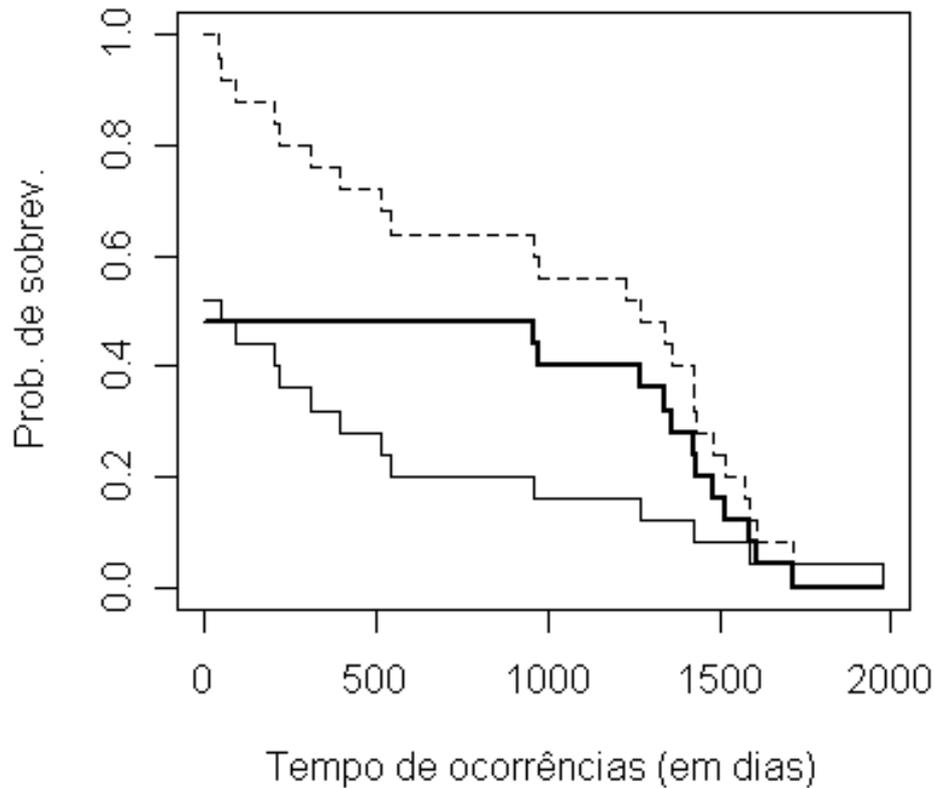


Figura 3.2 Estimativas das curvas da função conjunta de sobrevivência que é representada por (---), da função de subsobrevivência, devido ao evento específico, que é representada por (—) e da função de subsobrevivência, devido a outras causas, que é representada por (—), para os dados reais de transplantes renais.

Observamos que as curvas estimadas são monótonas decrescentes, descontínuas nos pontos onde ocorrem os eventos, formando patamares nos subintervalos entre cada dois pontos de "salto" consecutivos. Essa é uma característica muito comum na forma das estimativas das curvas de sobrevivência, obtidas pela aplicação da maioria dos estimadores conhecidos para este fim.

Tabela 3.2 Estimativas das funções empíricas e de subsobrevivência para os dados de transplantes renais.

$t$	$\hat{S}n(t)$	$\hat{S}e^*(t)$	$\hat{S}c^*(t)$	$t$	$\hat{S}n(t)$	$\hat{S}e^*(t)$	$\hat{S}c^*(t)$
0	1.000	0.520	0.480	1272 <sup>c</sup>	0.480	0.360	0.120
40	0.960	0.480	0.480	1343 <sup>c</sup>	0.440	0.320	0.120
43	0.920	0.440	0.480	1362 <sup>c</sup>	0.400	0.280	0.120
89	0.880	0.400	0.480	1424	0.360	0.280	0.080
201	0.840	0.360	0.480	1426 <sup>c</sup>	0.320	0.240	0.080
215	0.800	0.320	0.480	1434 <sup>c</sup>	0.280	0.200	0.080
308	0.760	0.280	0.480	1483 <sup>c</sup>	0.240	0.160	0.080
396	0.720	0.240	0.480	1518 <sup>c</sup>	0.200	0.120	0.080
512	0.680	0.200	0.480	1570	0.160	0.120	0.040
542	0.640	0.160	0.480	1588 <sup>c</sup>	0.120	0.080	0.040
956 <sup>c</sup>	0.600	0.160	0.440	1609 <sup>c</sup>	0.080	0.040	0.040
970 <sup>c</sup>	0.560	0.160	0.400	1713 <sup>c</sup>	0.040	0.000	0.040
1227	0.520	0.120	0.400	1973	0.000	0.000	0.000

### 3.6 O estimador atuarial adaptado para pequenas amostras

Este estimador da função de sobrevivência é conhecido na literatura como estimador atuarial das tábuas de vida, por Miller (1981). É muito utilizado, principalmente em estudos populacionais de grande escala, isto é, aqueles que envolvem uma quantidade de dados relativamente grande. Nestes casos, os dados apresentam-se, geralmente, agrupados arbitrariamente em subintervalos  $I_x = (x - \Delta_x, x)$ , onde  $x > 0$ , e não existe a possibilidade de se identificar a ordem em que ocorreram cada uma das falhas e/ou cada uma das censuras dentro de cada  $I_x$ . Para operacionalizá-lo para pequenas amostras, por Benze (1997), fazemos a ordenação (3.1) dos tempos de ocorrência dos eventos.

Então, a expressão principal deste estimador, adaptado para pequenas amostras

é dada, na notação da seção 3.1, por,

$$\hat{S}^{ea}(t) = \prod_{te_m \leq t} \left( 1 - \frac{M(te_m)}{V(te_m^-) + \frac{W(I_m)}{2}} \right), \quad (3.6)$$

em que  $\hat{S}^{ea}(t)$  é a estimativa da probabilidade de sobrevivência no tempo  $t$ , considerando o indivíduo exposto somente ao risco específico de interesse  $Re$  (com  $ea$  indicando que o estimador é o atuarial) e como limites dos subintervalos  $I_x$  os pontos de falha devido à causa  $Ce$ , isto é  $I_x \equiv I_m$ .

Para cada tempo  $t$  fixado, no intervalo  $(0, T_N)$ , realizamos o produto acumulado dos termos:  $\left( 1 - \frac{M(te_m)}{V(te_m^-) + \frac{W(I_m)}{2}} \right)$ , calculados em todos os pontos  $te_m$  de ocorrências de falhas que sejam anteriores ou no máximo igual ao ponto  $t$ .

As probabilidades de sobrevivência calculadas por este estimador não dependem dos tempos de duração censurantes mas das contagens  $W(I_m)$ , que estão implicitamente computadas em  $V(te_m^-)$  e explicitamente em  $\frac{W(I_m)}{2}$ . Note que em cada subintervalo  $I_m$ , metade das censuras (isto é  $\frac{W(I_m)}{2}$ ) são consideradas como indivíduos ainda em risco, imediatamente antes de  $te_m$ .

Para ilustrarmos o cálculo desse estimador utilizamos os dados de transplantes de rim do exemplo citado na seção 3.2. As estimativas da curva de sobrevivência pelo método atuarial são ilustradas na Figura 3.3 e os cálculos destas estimativas são apresentados na Tabela 3.3.

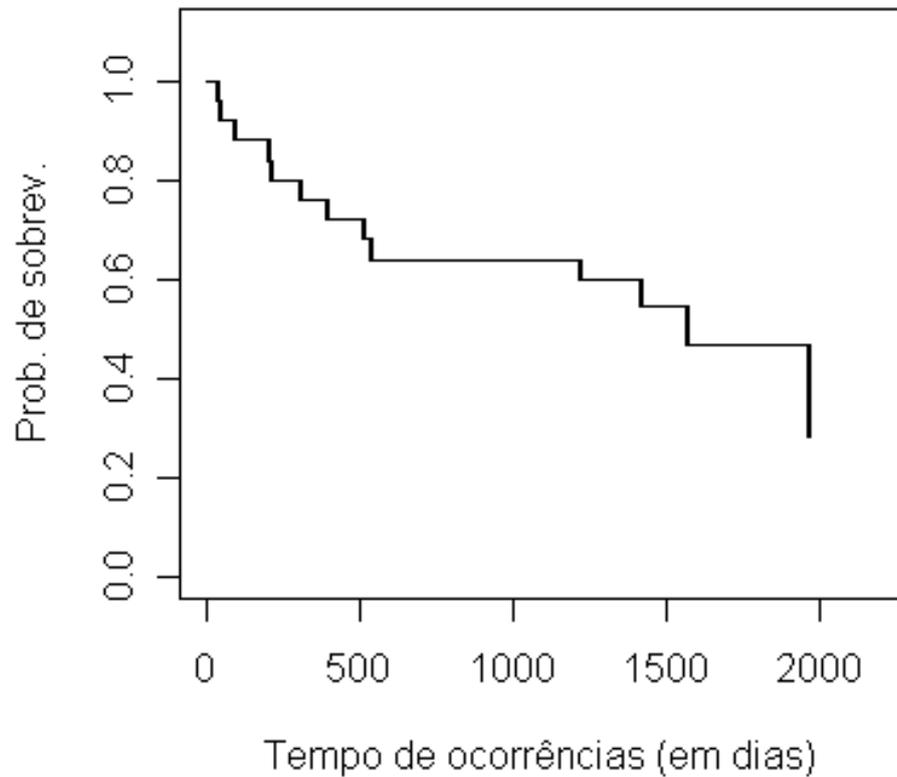


Figura 3.3 Estimativa da curva de sobrevivência  $ea$  para os dados reais de transplante renais.

Observamos que as curvas estimadas são monótonas decrescentes, descontínuas nos pontos onde ocorrem as falhas, formando patamares nos subintervalos entre cada dois pontos de "salto" correspondentes a estas falhas. É que, para este conjunto de dados, a cauda superior da curva não atinge o eixo das abscissas.

### 3.7 O estimador de Kaplan-Meier

Proposto inicialmente por Bôlmer (1912), o estimador produto-limite (P.L.) teve suas propriedades estudadas depois por Kaplan e Meier (1958), sendo por isso conhecido também como estimador de Kaplan-Meier. Para operacionalizá-lo devemos fazer a ordenação (3.1) dos tempos de ocorrência dos eventos.

A expressão principal deste estimador é estabelecida por,

$$\hat{S}e^{km}(t) = \prod_{te_m \leq t} \left( 1 - \frac{M(te_m)}{V(te_m^-)} \right). \quad (3.7)$$

Da mesma forma como no estimador atuarial adaptado, para cada tempo  $t$  fixado, no intervalo  $(0, T_N)$ , realizamos o produto acumulado destes termos:  $\left( 1 - \frac{M(te_m)}{V(te_m^-)} \right)$ , calculados em todos os pontos  $te_m$  de ocorrências de falhas que sejam anteriores ou no máximo igual ao ponto  $t$ .

As probabilidades de sobrevivência calculadas por este estimador não dependem diretamente dos tempos de duração censurantes mas das contagens  $W(I_m)$ , que estão implicitamente computadas em  $V(te_m^-)$ . Desta forma, podemos observar a manutenção mais prolongada, na curva estimada, de patamares elevados nos subintervalos onde encontram-se as censuras. Assim, analogamente ao que aconteceu ao estimador atuarial podemos observar a curva estimada assumindo a forma de patamares, nos subintervalos entre dois tempos consecutivos de ocorrências de falhas e para o mesmo conjunto de dados, a cauda superior da curva alcança o eixo das abscissas.

As estimativas das probabilidades de sobrevivência, obtidas através do estimador de Kaplan-Meier, para os dados de transplantes de rim, são apresentados na Tabela 3.3.

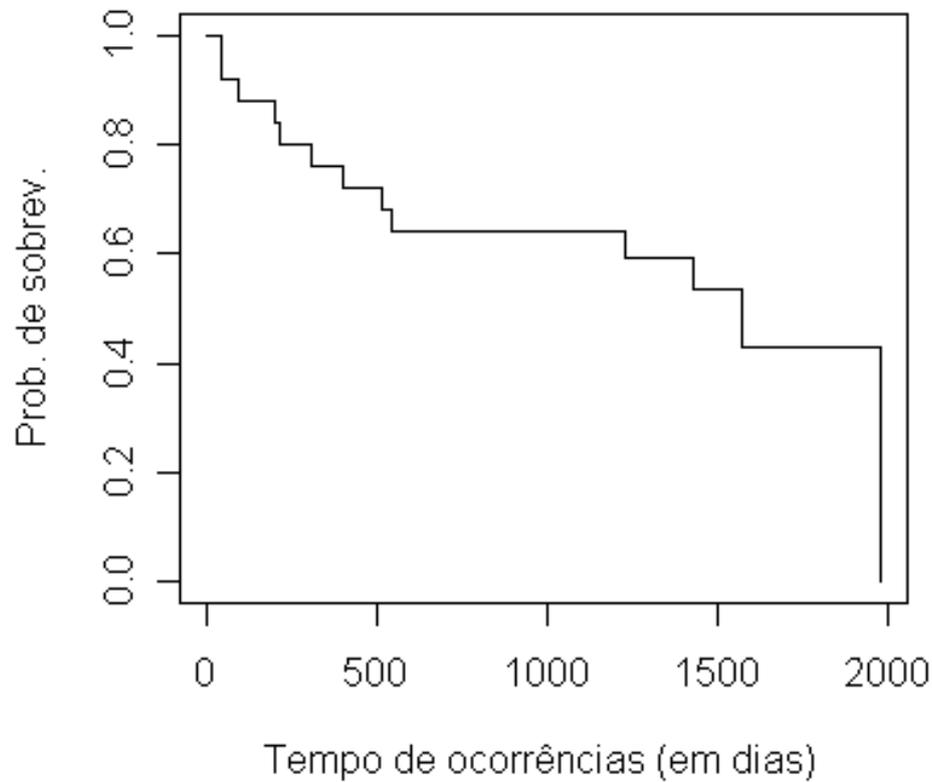


Figura.3.4 Estimativa da curva de sobrevivência *km* para os dados reais de transplantes renais.

A estimativa da correspondente função de risco acumulado pode ser obtida, por (2.6), calculando-se

$$\hat{\Lambda}^{xx}(t) = -\ln\{\hat{S}e^{xx}(t)\}, 0 < t < \infty, \quad (3.8)$$

onde ( $^{xx}$ ) correspondem às duas letras da sigla do estimador a qual estejamos nos referindo: *ea* para o estimador atuarial; *km* para o estimador de Kaplan-Meier; *en* para o estimador de Nelson-Aalen, o qual será comentado adiante, assim como, *ek* será usada para o estimador de Kitchin e *mk* para o estimador proposto modificado de Kitchin.

### 3.8 O estimador de Nelson

O estimador que apresentaremos agora, é muito empregado em estudos de experimentos industriais, da confiabilidade de peças e equipamentos. O estimador de Nelson da função de sobrevivência é dado por

$$\hat{S}e^{en}(t) = \exp \left( - \sum_{te_m \leq t} \frac{M(te_m)}{V(te_m^-)} \right). \quad (3.9)$$

Nota-se que o estimador é uma potência de base  $e$ , com o expoente dado por menos a soma acumulada. Esta soma acumulada é um estimador da função de risco acumulado, definido em (2.6), ou seja,

$$Re^{en}(t) = \sum_{te_m \leq t} \frac{M(te_m)}{V(te_m^-)}. \quad (3.10)$$

Na Tabela 3.3, a seguir, ilustramos as estimativas da curva de sobrevivência do estimador de Nelson, para os dados apresentados na seção 3.2 e a correspondente curva estimada de sobrevivência é apresentada na Figura 3.5, a seguir:

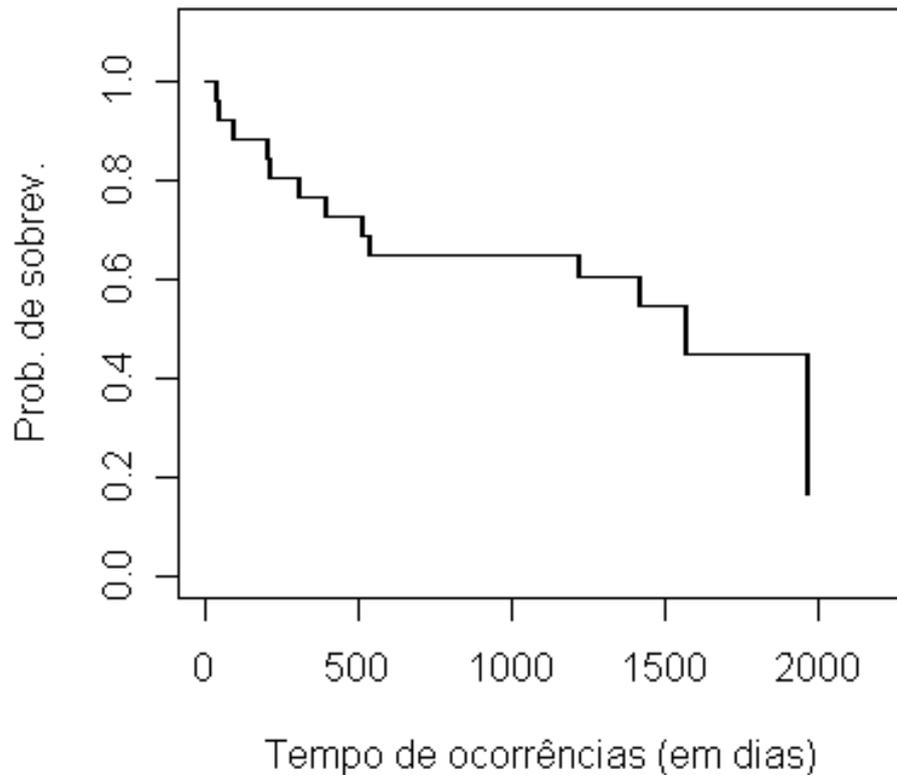


Figura 3.5 Estimativa da curva de sobrevivência *en* para os dados reais de transplantes renais.

Com relação a cauda superior o comportamento é análogo ao do estimador atuarial.

### 3.9 O estimador exponencial por partes de Kitchin

O estimador exponencial por partes de Kitchin não tem sido muito encontrado em aplicações tanto na área de saúde como na área de confiabilidade. Este estimador apresenta basicamente a mesma forma funcional do estimador *en*, no sentido da determinação da estimativa da curva de sobrevivência via estimativa da função risco acumulado. Entretanto, o estimador do risco é matematicamente mais elaborado do que aquele. O estimador ajusta funções exponenciais a cada dois pontos consecutivos de falhas, ao invés dos simples patamares de antes. Por isso ele é identificado pela sigla "PEXE" ("Piecwise

Exponential Estimator").

Essa maior "sofisticação" matemática do estimador de Kitchin, quando comparado com o de Nelson ou o de Kaplan-Meier, exige do usuário algum esforço adicional para a sua compreensão e operacionalização. Esse deve ser um dos principais motivos para a sua pouca divulgação e aplicação em problemas reais.

Uma das suas limitações é que o estimador inicialmente elaborado por Kitchin em 1980 não admite empates. Kim e Proschan (1991) usam a mesma versão do estimador de Kitchin, com a única diferença incluir empates nos cálculos.

A expressão geral do estimador de Kitchin é dada por

$$\hat{S}e^{ek}(t) = \exp \left( -(te - te_{m-1})\hat{\lambda}e_m^{ek} - \sum_{te_m \leq t} [(te_{m-1} - te_{m-2})\hat{\lambda}e_{m-1}^{ek}] \right) \quad (3.11)$$

onde as taxas instantâneas de falha são estimadas por

$$\begin{aligned} \hat{\lambda}e_m^{ek} &= \frac{1}{A_{ek}} & (3.12) \\ &= \text{inverso da soma das áreas sob a curva determinada pela} \\ &\quad \text{curva empírica } \hat{S}n \text{ (curva de contagem das falhas e censuras)} \\ &\quad \text{entre } te_{m-1} \text{ e } te, \end{aligned}$$

em que as letras  $ek$  são aqui empregadas como sigla do estimador de Kitchin (1980).

Para este estimador, as estimativas das probabilidades de sobrevivência ao risco  $Re$  decrescem em todos os pontos do subintervalo  $I_m$ , inclusive nos tempos de censuras, aumentando a intensidade de decrescimento na "passagem" de um subintervalo para o próximo no ponto  $te_m$ . Quando  $t$  ultrapassar o último tempo de falha da amostra, a estimativa da probabilidade de sobrevivência ao risco  $Re$ , nestes tempos  $t$  usará a mesma taxa determinada no último tempo de falha (o último pedaço exponencial é prolongado até a posição do ponto  $t$ ).

**Observação 2:** Para o estimador de Kitchin, esta pode ser uma possível solução para o problema (ii) a indefinição no final da curva de sobrevivência estimada, quando a última unidade a abandonar o estudo é uma censura.

Apresentamos na Figura 3.6 a estimativa da correspondente curva de sobrevivência, para o conjunto de dados, citados na seção 3.2.

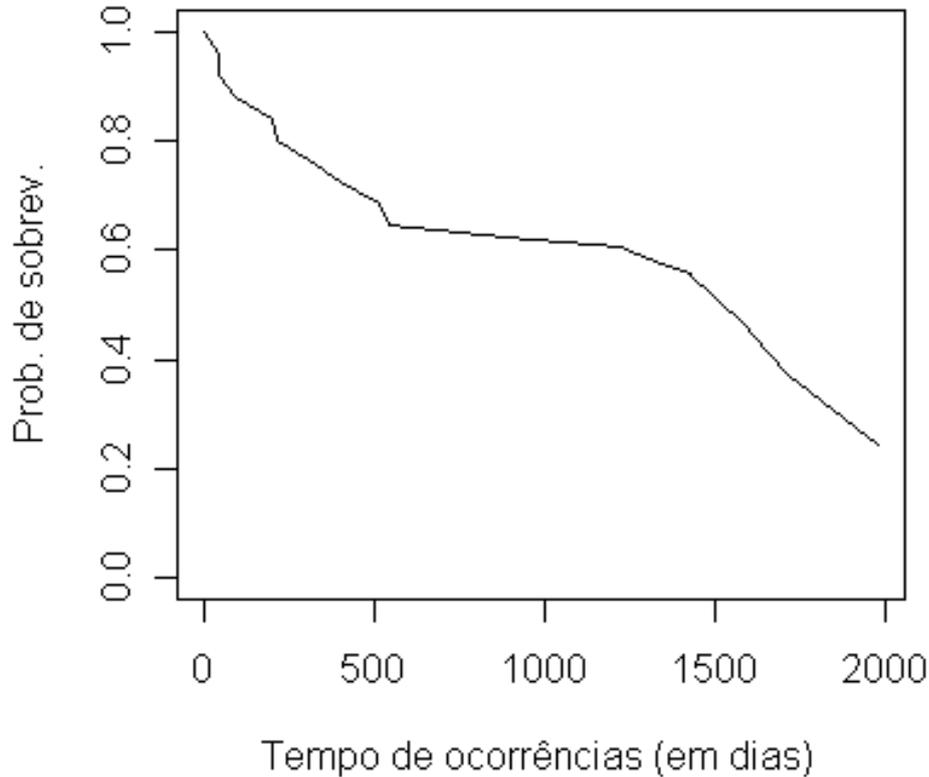


Figura 3.6 Estimativa da curva de sobrevivência  $ek$  para os dados reais de transplantes renais.

Observamos que para este conjunto de dados, a cauda superior da curva também não atinge o eixo das abscissas.

A função de sobrevivência deste estimador é constituída por uma curva contínua, nos subintervalos  $t_{m-1} < t < t_m$ , por partes  $s_m(t)$  de funções exponenciais dadas por

$$s_m(t) = \exp(-[(te_{m-1})\Lambda e^{ek} + (t - te_{m-1})\lambda e_m^{ek}]) \quad (3.13)$$

O primeiro termo dentro do expoente, corresponde ao risco acumulado até o limite inferior no intervalo  $I_m$  considerado.

A determinação da área sob a curva de Kitchin faz-se com a integração de cada

um destes pedaços de funções exponenciais, entre os extremos dos correspondentes subintervalos e a soma dos resultados, ou seja,

$$\begin{aligned}
 \text{Área sob a curva de Kitchin} &= \sum_{m=1}^M \int_{t_{m-1}}^{t_m} s_m(t) dt & (3.14) \\
 &= \sum_{m=1}^M \exp \left\{ -\Lambda e^{ek}(t_{m-1}) + \lambda e_m^{ek} * t_{m-1} \right\} * \\
 &\quad * \left[ \left( \frac{1}{\lambda e_m^{ek}} \right) \left( \exp \left\{ -\lambda e_m^{ek} * t_{m-1} \right\} - \exp \left\{ -\lambda e_m^{ek} * t_m \right\} \right) \right].
 \end{aligned}$$

### 3.10 Comparação das estimativas da curva de sobrevivência, segundo os procedimentos não paramétricos apresentados

Nesta seção, comparamos as estimativas da curva de sobrevivência, obtidas pelos estimadores apresentados. Na Figura 3.7 representamos os valores registrados na Tabela 3.3 e verificamos a posição das estimativas obtidas da curva de sobrevivência em relação aos correspondentes limites de Peterson, com *LS* representando o limite superior e *LI* o limite inferior, para os dados de trasplantes renais.

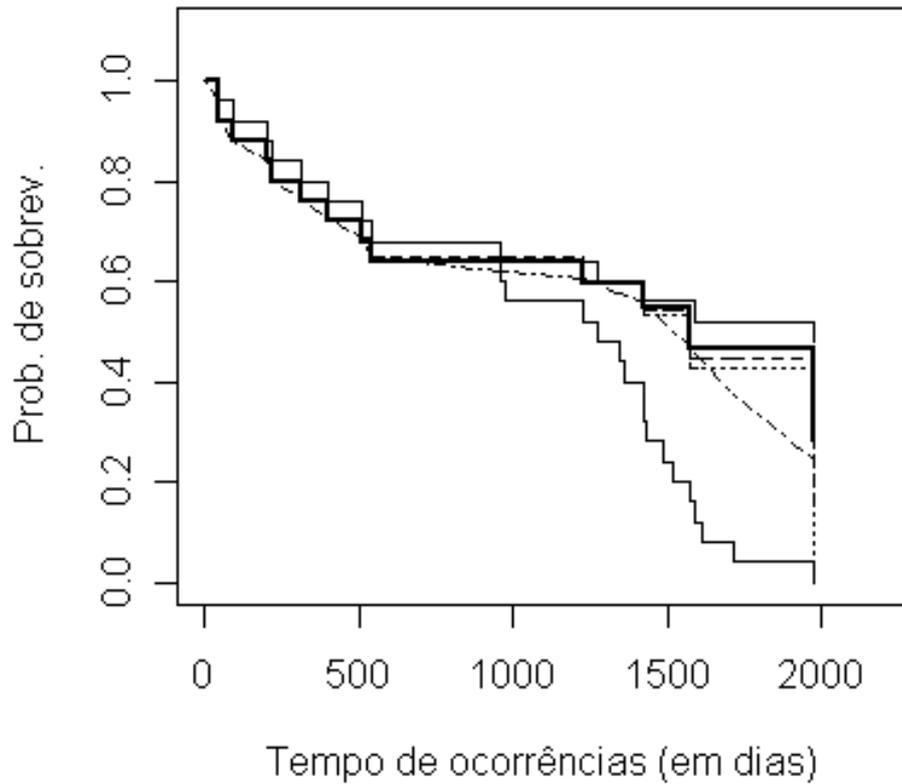


Figura 3.7 Comparação das estimativas  $ea$ ,  $km$ ,  $en$  e  $ek$ , e situação delas em relação aos dados de transplantes renais. O estimador  $ea$  é denotado por (—), o  $en$  é denotado por (---), o  $km$  é denotado por (....) o  $ek$  é denotado por (- - - -) e os limites de Peterson são denotados por (—).

Observamos que todas as curvas estimadas estão dentro dos limites de Peterson. O estimador  $ek$ , comparado com os demais, subestima as probabilidades de sobrevivência e podemos observar que a curva  $ek$  apresenta-se localizada no meio da faixa determinada pelos limites de Peterson. Devemos ressaltar que os limites de Peterson (1976) estabelecem faixas da localização da verdadeira curva de sobrevivência, sem a suposição ou não de independência entre os tempos de falha e de censura. Mas essas faixas se alargam muito com o passar do tempo, tornando-se inutilizáveis. Os limites de Link (1989), sugerem uma faixa mais estreita, porém dependem de estimativas em forma de

escada, isto é, os limites superiores, estabelecidos na suposição de total independência, são determinados pelas estimativas de Kaplan-Meier e os limites inferiores, estabelecidos na suposição de total dependência, são determinados pelas estimativas empíricas (estimador empírico). Dentro desse último raciocínio, o nosso trabalho, é um ensaio no sentido de melhorar essa faixa, sugerindo como limite superior as estimativas de Kitchin e como limite inferior as estimativas do estimador proposto modificado de Kitchin ( $mk$ ), que apresentamos no Capítulo 5. Pois as estimativas  $ek$  são contínuas e pressupõe independência, e as estimativas  $mk$  são contínuas e apresentam-se mais próximas das estimativas empíricas.

Tabela 3.3 Estimativas  $ea$ ,  $km$ ,  $en$ ,  $ek$  e limites inferior ( $LI$ ) e superior ( $LS$ ) de Peterson da curva de sobrevivência para os dados de transplantes renais.

$t$	$ea$	$km$	$en$	$ek$	$LI$	$LS$
0	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.961	0.921	0.920	0.920
89	0.880	0.880	0.882	0.882	0.880	0.880
201	0.840	0.840	0.843	0.843	0.840	0.840
215	0.800	0.800	0.803	0.803	0.800	0.800
308	0.760	0.760	0.764	0.764	0.760	0.760
396	0.720	0.720	0.725	0.725	0.720	0.720
512	0.680	0.680	0.686	0.686	0.680	0.680
542	0.640	0.640	0.647	0.647	0.640	0.640
956 <sup>c</sup>	0.640	0.640	0.647	0.621	0.600	0.640
970 <sup>c</sup>	0.640	0.640	0.647	0.621	0.560	0.640
1227	0.597	0.594	0.602	0.606	0.520	0.600
1272 <sup>c</sup>	0.597	0.594	0.602	0.594	0.480	0.600
1343 <sup>c</sup>	0.597	0.594	0.602	0.575	0.440	0.600
1362 <sup>c</sup>	0.597	0.594	0.602	0.570	0.400	0.600
1424	0.545	0.534	0.545	0.555	0.360	0.560
1426 <sup>c</sup>	0.545	0.534	0.5452	0.554	0.320	0.560
1434 <sup>c</sup>	0.545	0.534	0.545	0.549	0.280	0.560
1483 <sup>c</sup>	0.545	0.534	0.545	0.520	0.240	0.560
1518 <sup>c</sup>	0.545	0.534	0.545	0.500	0.200	0.560

Tabela 3.3 (continuação) Estimativas  $ea$ ,  $km$ ,  $en$ ,  $ek$  e limites inferior ( $LI$ ) e superior ( $LS$ ) de Peterson da curva de sobrevivência para os dados de transplantes renais.

$t$	$ea$	$km$	$en$	$ek$	$LI$	$LS$
1570	0.467	0.427	0.446	0.471	0.160	0.520
1588 <sup>c</sup>	0.467	0.427	0.446	0.458	0.120	0.520
1609 <sup>c</sup>	0.467	0.427	0.446	0.442	0.080	0.520
1713 <sup>c</sup>	0.467	0.427	0.446	0.372	0.040	0.520
1973	0.280	0.000	0.164	0.241	0.000	0.480

No Capítulo seguinte apresentamos os problemas que prioritariamente nos propomos a estudar nesta dissertação. Apresentamos os efeitos das perturbações Tipo I e II em conjuntos de dados reais, primeiramente no exemplo com dados clínicos de transplantes renais humanos, no qual realizamos uma única perturbação e depois duas perturbações. Um outro exemplo com dados reais de transplantes de medula óssea também é apresentado neste Capítulo, onde realizamos uma única perturbação no conjunto de dados.

# Capítulo 4

## Alguns Problemas de Interesse

Neste Capítulo apresentamos alguns problemas que estudamos. Ilustramos em dois exemplos reais os efeitos das perturbações, com os dados clínicos reais de transplantes renais humanos e transplantes de medula óssea, no qual para estas perturbações fizemos modificações específicas. Comparamos as estimativas dos dados reais, na situação normal, comparados com os dados sujeitos às perturbação do Tipo I e II (indicada pelo símbolo  $\delta$ ).

Dentre os problemas que citamos na introdução, abordamos prioritariamente os seguintes: ix) qual estimador apresenta melhor desempenho dentre os mais utilizados; v) a vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier, e muito provavelmente dos demais, a pequenas perturbações provocadas no conjunto de dados e propor um novo estimador não paramétrico para a curva de sobrevivência, sob dados censurados.

Oakes (1993) cita em seu artigo, que o estimador de Kaplan Meier sofre alterações em suas estimativas quando os dados são perturbados. Na perturbação realizada por Oakes (1993), ele aponta que as estimativas  $km$  se elevam nas vizinhanças do tempo de perturbação e depois decrescem em relação a situação normal (de não perturbação), até o final dos tempos observados.

Pretendemos identificar via simulação bootstrap, dentre os estimadores já consolidados, aquele(s) que seja(m) menos vulnerável(eis) ao problema v). Entretanto, iniciamos apontando primeiramente um exemplo real em que podemos verificar os efeitos das perturbações Tipo I e II. O que é apresentado nas Tabelas 4.1 e 4.2, com os dados clínicos

reais de transplantes renais humanos, no qual para a perturbação do Tipo I (indicada pelo símbolo ') é feita uma modificação específica de ocorrência de falha no tempo igual a 512 e, alternativamente, fixada para um tempo adiante do primeiro tempo censurante 956\*. Para a perturbação do Tipo II, na ocorrência de falha do tempo 512 assumimos que o mesmo é um tempo de censura. Podemos observar que, para este conjunto de dados, o efeito da perturbação, sobre a estimativa da curva de sobrevivência, ocorre não só para o estimador  $km$  mas também para os demais aqui estudados. Neste exemplo, observamos com os dados de transplantes renais que o "estimador  $ek$  é aparentemente mais resistente à perturbação do Tipo I", enquanto que o "estimador  $mk$  é aparentemente mais resistente à perturbação do Tipo II" que os outros estimadores aqui apresentados (Veja Tabelas 4.1 e 4.2 ). Salientamos que a resistência às perturbações consideradas são, em princípio, constatações empíricas e baseadas nas diferenças relativas entre os estimadores obtidos com os dados originais e com a perturbação.

Observação: Nas Tabelas 4.1, 4.2, 4.3 e 4.4 ressaltamos com um ponto (.) os resultados das estimativas não existentes, para todos os estimadores, na situação normal (de não perturbação) e nas situações com as perturbações.

Tabela 4.1 Estimativas dos dados não perturbados comparados com os dados sob o efeito de 1 perturbação no  $t$  igual a 512, para a perturbação do Tipo I.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.920	0.920	0.961	0.921	0.921	0.921	0.921	0.921
89	0.880	0.880	0.880	0.880	0.882	0.882	0.882	0.882	0.882	0.882
201	0.840	0.840	0.840	0.840	0.843	0.843	0.843	0.843	0.843	0.843
215	0.800	0.800	0.800	0.800	0.803	0.803	0.803	0.803	0.803	0.803
308	0.760	0.760	0.760	0.760	0.764	0.764	0.764	0.764	0.764	0.764
396	0.720	0.720	0.720	0.720	0.725	0.725	0.725	0.725	0.725	0.725
512	0.680	.	0.680	.	0.686	.	0.686	.	0.686	.
542	0.640	0.720	0.640	0.720	0.647	0.725	0.647	0.694	0.647	0.686
956 <sup>c</sup>	0.640	0.680	0.640	0.680	0.647	0.686	0.621	0.686	0.603	0.649
512'	.	0.680	.	0.680	.	0.686	.	0.648	.	0.611
970 <sup>c</sup>	0.640	0.640	0.640	0.634	0.647	0.642	0.621	0.647	0.563	0.575
1227	0.597	0.595	0.594	0.589	0.602	0.597	0.606	0.602	0.526	0.536
1272 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.594	0.590	0.479	0.488
1343 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.575	0.572	0.436	0.444
1362 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.570	0.567	0.397	0.405
1424	0.545	0.544	0.534	0.530	0.545	0.540	0.555	0.552	0.361	0.368
1426 <sup>c</sup>	0.545	0.544	0.534	0.530	0.545	0.540	0.554	0.551	0.304	0.310
1434 <sup>c</sup>	0.545	0.544	0.534	0.530	0.545	0.540	0.549	0.546	0.256	0.261
1483 <sup>c</sup>	0.545	0.544	0.534	0.530	0.545	0.540	0.520	0.517	0.216	0.220
1518 <sup>c</sup>	0.545	0.544	0.534	0.530	0.545	0.540	0.500	0.497	0.182	0.185

Tabela 4.1 (continuação) Estimativas dos dados não perturbados comparados com os dados sob o efeito de 1 perturbação no  $t$  igual a 512, para a perturbação do Tipo I.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
1518 <sup>c</sup>	0.545	0.544	0.534	0.530	0.545	0.540	0.500	0.497	0.182	0.185
1588 <sup>c</sup>	0.467	0.466	0.427	0.424	0.446	0.442	0.458	0.455	0.076	0.077
1609 <sup>c</sup>	0.467	0.466	0.427	0.424	0.446	0.442	0.442	0.439	0.037	0.038
1713 <sup>c</sup>	0.467	0.466	0.427	0.424	0.446	0.442	0.372	0.370	0.018	0.019
1973	0.280	0.279	0.000	0.000	0.164	0.162	0.241	0.240	0.009	0.009

Tabela 4.2 Estimativas dos dados não perturbados comparados com os dados sob o efeito de 1 perturbação no  $t$  igual a 512, para a perturbação do Tipo II.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.920	0.920	0.961	0.921	0.921	0.921	0.921	0.921
89	0.880	0.880	0.880	0.880	0.882	0.882	0.882	0.882	0.882	0.882
201	0.840	0.840	0.840	0.840	0.843	0.843	0.843	0.843	0.843	0.843
215	0.800	0.800	0.800	0.800	0.803	0.803	0.803	0.803	0.803	0.803
308	0.760	0.760	0.760	0.760	0.764	0.764	0.764	0.764	0.764	0.764
396	0.720	0.720	0.720	0.720	0.725	0.725	0.725	0.725	0.725	0.725
512	0.680	.	0.680	.	0.686	.	0.686	.	0.683	.
512 <sup>c</sup>	.	0.720	.	0.720	.	0.725	.	0.693	.	0.684
542	0.640	0.678	0.640	0.677	0.647	0.684	0.647	0.685	0.647	0.646
956 <sup>c</sup>	0.640	0.678	0.640	0.677	0.647	0.684	0.621	0.659	0.603	0.603
970 <sup>c</sup>	0.640	0.678	0.640	0.677	0.647	0.684	0.621	0.658	0.563	0.562
1227	0.597	0.633	0.594	0.629	0.602	0.636	0.606	0.642	0.526	0.525
1272 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.594	0.629	0.479	0.478
1343 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.575	0.610	0.436	0.435
1362 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.570	0.605	0.397	0.396
1424	0.545	0.578	0.534	0.566	0.545	0.576	0.555	0.588	0.361	0.361
1426 <sup>c</sup>	0.545	0.578	0.534	0.566	0.545	0.576	0.554	0.587	0.304	0.304
1434 <sup>c</sup>	0.545	0.578	0.534	0.566	0.545	0.576	0.549	0.582	0.256	0.256
1483 <sup>c</sup>	0.545	0.578	0.534	0.566	0.545	0.576	0.520	0.551	0.216	0.216
1518 <sup>c</sup>	0.545	0.578	0.534	0.566	0.545	0.576	0.500	0.530	0.182	0.182

Tabela 4.2 (continuação) Estimativas dos dados não perturbados comparados com os dados sob o efeito de 1 perturbação no  $t$  igual a 512, para a perturbação do Tipo II.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
1570	0.467	0.495	0.427	0.453	0.446	0.471	0.471	0.500	0.153	0.153
1588 <sup>c</sup>	0.467	0.495	0.427	0.453	0.446	0.471	0.458	0.485	0.076	0.076
1609 <sup>c</sup>	0.467	0.495	0.427	0.453	0.446	0.471	0.442	0.468	0.037	0.037
1713 <sup>c</sup>	0.467	0.495	0.427	0.453	0.446	0.471	0.372	0.394	0.018	0.018
1973	0.280	0.297	0.000	0.000	0.164	0.173	0.241	0.256	0.009	0.009

A análise da Tabela 4.1 nos mostra que o  $ea$  foi o mais resistente, para uma única perturbação do Tipo I. O segundo melhor foi o  $ek$  e o menos resistente foi o  $mk$ . Conforme observamos, para este conjunto específico de dados, o efeito da perturbação do Tipo I, sobre a estimativa da curva de sobrevivência, ocorre também para os estimadores  $ea$ ,  $km$ ,  $en$ ,  $ek$  e  $mk$ . Podemos ver, claramente, que antes e no tempo de perturbação a estimativa da curva de sobrevivência se eleva em relação a situação normal (de não perturbação), para todos os estimadores, porém para o  $mk$  as estimativas se elevam nas vizinhanças do tempo de perturbação e continuam elevadas até o final dos tempos.

A Tabela 4.2 apresenta o efeito de uma única perturbação do Tipo II e a análise destes efeitos nos revela que o  $mk$  foi o mais resistente. O segundo mais resistente foi o  $en$  e o  $ea$  foi o menos resistente. Podemos observar, que antes e no tempo de perturbação a estimativa da curva de sobrevivência se eleva, em relação a situação normal (de não perturbação) e continuam elevadas até o final dos tempos, para os estimadores  $ea$ ,  $km$ ,  $en$  e  $ek$ , porém notamos que o efeito da perturbação para as estimativas  $mk$  ocorre somente nas vizinhanças do tempo de perturbação, e ao contrário dos demais estimadores, a estimativa da curva de sobrevivência diminui suavemente, em relação a situação normal (de não perturbação) mantendo-se depois, exatamente, as mesmas estimativas anteriores à perturbação.

Nas Tabelas 4.3 e 4.4 apresentamos os resultados do efeito de duas perturbações

Tipo I e II, para o mesmo conjunto de dados. Além da perturbação do Tipo I (indicada pelo símbolo ') feita com a modificação específica de ocorrência de falha no tempo igual a 512 e, alternativamente, fixada para um tempo adiante do primeiro tempo censurante 956<sup>c</sup>, acrescentamos mais uma modificação de falha no tempo igual a 1424 para adiante do primeiro tempo censurante 1426<sup>c</sup>. Para a perturbação do Tipo II, além da já efetuada alternativamente, no tempo de falha 512 para 512<sup>c</sup> incluímos uma segunda modificação no tempo 1424 para 1424<sup>c</sup>.

Tabela 4.3 Estimativas dos dados não perturbados comparados com os dados sob o efeito de duas perturbações no  $t$  igual a 512 e 1424, para a perturbação do Tipo I.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.920	0.920	0.961	0.921	0.921	0.921	0.921	0.921
89	0.880	0.880	0.880	0.880	0.882	0.882	0.882	0.882	0.882	0.882
201	0.840	0.840	0.840	0.840	0.843	0.843	0.843	0.843	0.843	0.843
215	0.800	0.800	0.800	0.800	0.803	0.803	0.803	0.803	0.803	0.803
308	0.760	0.760	0.760	0.760	0.764	0.764	0.764	0.764	0.764	0.764
396	0.720	0.720	0.720	0.720	0.725	0.725	0.725	0.725	0.725	0.725
512	0.680	.	0.680	.	0.686	.	0.686	.	0.686	.
542	0.640	0.720	0.640	0.720	0.647	0.725	0.647	0.694	0.647	0.686
956 <sup>c</sup>	0.640	0.680	0.640	0.680	0.647	0.686	0.621	0.686	0.603	0.649
512'	.	0.680	.	0.680	.	0.686	.	0.648	.	0.611
970 <sup>c</sup>	0.640	0.640	0.640	0.634	0.647	0.642	0.621	0.647	0.563	0.575
1227	0.597	0.595	0.594	0.589	0.602	0.597	0.606	0.602	0.526	0.536
1272 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.594	0.591	0.479	0.485
1343 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.575	0.573	0.436	0.439
1362 <sup>c</sup>	0.597	0.595	0.594	0.589	0.602	0.597	0.570	0.568	0.397	0.398
1424	0.545	.	0.534	.	0.545	.	0.555	.	0.361	.
1426 <sup>c</sup>	0.545	0.595	0.534	0.589	0.545	0.597	0.554	0.554	0.304	0.360
1424'	.	0.595	.	0.589	.	0.597	.	0.553	.	0.326
1434 <sup>c</sup>	0.545	0.544	0.534	0.515	0.545	0.527	0.549	0.551	0.256	0.295
1483 <sup>c</sup>	0.545	0.544	0.534	0.515	0.545	0.527	0.520	0.519	0.216	0.247
1518 <sup>c</sup>	0.545	0.544	0.534	0.515	0.545	0.527	0.500	0.497	0.182	0.207

Tabela 4.3 (continuação) Estimativas dos dados não perturbados comparados com os dados com os dados sob o efeito de duas perturbações no  $t$  igual a 512 e 1424, para a perturbação do Tipo I.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
1588 <sup>c</sup>	0.467	0.460	0.427	0.412	0.446	0.431	0.458	0.453	0.076	0.086
1609 <sup>c</sup>	0.467	0.460	0.427	0.412	0.446	0.431	0.442	0.437	0.037	0.042
1713 <sup>c</sup>	0.467	0.460	0.427	0.412	0.446	0.431	0.372	0.368	0.018	0.021
1973	0.280	0.276	0.000	0.000	0.164	0.158	0.241	0.239	0.009	0.010

Tabela 4.4 Estimativas dos dados não perturbados comparados sob o efeito de 2 perturbações no  $t$  igual a 512 e 1424, para a perturbação do Tipo II.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.920	0.920	0.961	0.921	0.921	0.921	0.921	0.921
89	0.880	0.880	0.880	0.880	0.882	0.882	0.882	0.882	0.882	0.882
201	0.840	0.840	0.840	0.840	0.843	0.843	0.843	0.843	0.843	0.843
215	0.800	0.800	0.800	0.800	0.803	0.803	0.803	0.803	0.803	0.803
308	0.760	0.760	0.760	0.760	0.764	0.764	0.764	0.764	0.764	0.764
396	0.720	0.720	0.720	0.720	0.725	0.725	0.725	0.725	0.725	0.725
512	0.680	.	0.680	.	0.686	.	0.686	.	0.683	.
512 <sup>c</sup>	.	0.720	.	0.720	.	0.725	.	0.693	.	0.684
542	0.640	0.678	0.640	0.677	0.647	0.684	0.647	0.685	0.647	0.646
956 <sup>c</sup>	0.640	0.678	0.640	0.677	0.647	0.684	0.621	0.659	0.603	0.603
970 <sup>c</sup>	0.640	0.678	0.640	0.677	0.647	0.684	0.621	0.658	0.563	0.562
1227	0.597	0.633	0.594	0.629	0.602	0.636	0.606	0.642	0.526	0.525
1272 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.594	0.633	0.479	0.449
1343 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.575	0.619	0.436	0.384
1362 <sup>c</sup>	0.597	0.633	0.594	0.629	0.602	0.636	0.570	0.615	0.397	0.329
1424	0.545	.	0.534	.	0.545	.	0.555	.	0.361	.
1424'	.	0.633	.	0.629	.	0.636	.	0.603	.	0.281
1426 <sup>c</sup>	0.545	0.633	0.534	0.629	0.545	0.636	0.554	0.601	0.304	0.241
1434 <sup>c</sup>	0.545	0.633	0.534	0.629	0.545	0.636	0.549	0.601	0.256	0.206
1483 <sup>c</sup>	0.545	0.633	0.534	0.629	0.545	0.636	0.520	0.592	0.216	0.176
1518 <sup>c</sup>	0.545	0.633	0.534	0.629	0.545	0.636	0.500	0.585	0.182	0.151
1570	0.467	0.563	0.427	0.503	0.446	0.521	0.471	0.576	0.139	0.129
1588 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.458	0.559	0.069	0.064
1609 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.442	0.540	0.035	0.031

Tabela 4.4 (continuação) Estimativas dos dados não perturbados comparados com os dados sob o efeito de 2 perturbações no  $t$  igual a 512 e 142, para a perturbação do Tipo II.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
1713 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.372	0.454	0.017	0.015
1973	0.280	0.337	0.000	0.000	0.164	0.191	0.241	0.295	0.009	0.007
1570	0.467	0.563	0.427	0.503	0.446	0.521	0.471	0.576	0.139	0.129
1588 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.458	0.559	0.069	0.064
1609 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.442	0.540	0.035	0.031
1713 <sup>c</sup>	0.467	0.563	0.427	0.503	0.446	0.521	0.372	0.454	0.017	0.015
1973	0.280	0.337	0.000	0.000	0.164	0.191	0.241	0.295	0.009	0.007

A Tabela 4.3 nos mostra que o  $ea$  continuou o mais resistente para duas perturbações do Tipo I. O segundo melhor continuou sendo o  $ek$  e o menos resistente o  $mk$ . O comportamento das estimativas também foram os mesmos em relação a uma perturbação. A análise da Tabela 4.4 também nos mostra o mesmo comportamento em relação a uma perturbação do Tipo II. Porém, o menos resistente foi o  $ek$ . As estimativas apresentaram-se também com o mesmo comportamento em relação a uma perturbação.

## 4.1 Outro exemplo com dados reais

Considere o conjunto de dados extraído de Avalos (1993), referente a 43 tempos de sobrevida, de pacientes que foram submetidos ao transplante de medula óssea. A Tabela 4.5, apresenta os tempos de sobrevida (em dias), para cada paciente da amostra estudada. Realizamos as mesmas perturbações do Tipo I e II. Observamos neste outro conjunto de dados que o estimador proposto  $mk$  apresentou-se aparentemente mais resistente às perturbações do Tipo I e II em relação aos demais estimadores estudados. (Veja Tabelas 4.6 e 4.7).

Podemos observar que, enquanto, para o conjunto de dados de transplantes renais o estimador de Kitchin ( $ek$ ) foi o que aparentemente apresentou menor alteração, para o conjunto de dados de transplantes de medula óssea o estimador proposto  $mk$  parece não sofrer alterações, isto é, foi aparentemente mais resistente às perturbações Tipo I e II.

Neste contexto, para confirmarmos a resistência do estimador proposto, consideramos um estudo de simulação que examina o desempenho dos estimadores já consolidados.

Mediante esses resultados obtidos para os dois conjuntos específicos de dados considerados, pretendemos identificar via simulação bootstrap, dentre os estimadores já consolidados, aquele(s) que seja(m) menos vulnerável(eis) ao problema destas perturbações.

Tabela 4.5. Tempos de sobrevivência (em dias) para cada paciente em estudo onde ( $^c$ ) representa os tempos censurados.

$t_1=2$	$t_2=4$	$t_3=28$	$t_4=30$	$t_5=32$
$t_6=36$	$t_7=41$	$t_8=42$	$t_9=49$	$t_{10}=52$
$t_{11}=53$	$t_{12}=57$	$t_{13}=62$	$t_{14}=63$	$t_{15}=72$
$t_{16}=77$	$t_{17}=79$	$t_{18}=81$	$t_{19}=81$	$t_{20}=84$
$t_{21}=108$	$t_{22}=132$	$t_{23}=140$	$t_{24}=180^c$	$t_{25}=210^c$
$t_{26}=252$	$t_{27}=307^c$	$t_{28}=357$	$t_{29}=406^c$	$t_{30}=446^c$
$t_{31}=476^c$	$t_{32}=484^c$	$t_{33}=524$	$t_{34}=748^c$	$t_{35}=933^c$
$t_{36}=1037^c$	$t_{37}=1078^c$	$t_{38}=1183^c$	$t_{39}=1290^c$	$t_{40}=1345^c$
$t_{41}=1560^c$	$t_{42}=2114^c$	$t_{43}=2144^c$		

Tabela 4.6 Estimativas dos dados não perturbados comparados com os dados perturbados no  $t$  igual a 108, para a perturbação do Tipo I, para os dados reais.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
2	0.976	0.976	0.976	0.976	0.977	0.977	0.977	0.977	0.977	0.977
4	0.953	0.953	0.953	0.953	0.954	0.954	0.954	0.954	0.954	0.954
28	0.930	0.930	0.930	0.930	0.931	0.931	0.931	0.931	0.931	0.931
30	0.907	0.907	0.907	0.907	0.908	0.908	0.908	0.908	0.908	0.908
32	0.883	0.883	0.883	0.883	0.885	0.885	0.885	0.885	0.885	0.885
36	0.860	0.860	0.860	0.860	0.862	0.862	0.862	0.862	0.862	0.862
41	0.837	0.837	0.837	0.837	0.8391	0.839	0.839	0.839	0.839	0.839
42	0.814	0.814	0.814	0.814	0.816	0.816	0.816	0.816	0.816	0.816
49	0.790	0.790	0.790	0.790	0.793	0.793	0.793	0.793	0.793	0.793
52	0.767	0.767	0.767	0.767	0.770	0.770	0.770	0.770	0.770	0.770
53	0.744	0.744	0.744	0.744	0.747	0.747	0.747	0.747	0.747	0.747
57	0.720	0.720	0.720	0.720	0.724	0.724	0.724	0.724	0.724	0.724
62	0.697	0.697	0.697	0.697	0.701	0.701	0.701	0.701	0.701	0.701
63	0.674	0.674	0.674	0.674	0.678	0.678	0.678	0.678	0.678	0.678
72	0.651	0.651	0.651	0.651	0.655	0.655	0.655	0.655	0.655	0.655
77	0.627	0.627	0.627	0.627	0.632	0.632	0.632	0.632	0.632	0.632
79	0.604	0.604	0.604	0.604	0.609	0.609	0.609	0.609	0.609	0.609
81	0.581	0.581	0.581	0.581	0.586	0.586	0.586	0.586	0.586	0.586
81	0.558	0.558	0.558	0.558	0.563	0.563	0.563	0.563	0.563	0.563
84	0.534	0.534	0.534	0.534	0.540	0.540	0.540	0.540	0.540	0.540
108	0.511	.	0.511	.	0.5113	.	0.5173	.	0.5173	.
132	0.488	0.534	0.488	0.534	0.494	0.540	0.494	0.517	0.494	0.517
140	0.465	0.511	0.465	0.511	0.471	0.517	0.471	0.494	0.471	0.494

Tabela 4.6 (continuação) Estimativas dos dados não perturbados comparados com os dados perturbados no  $t$  igual a 108, para a perturbação do Tipo I, para os dados reais.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
180	0.465	0.488	0.465	0.488	0.471	0.494	0.462	0.490	0.446	0.470
210	0.465	0.488	0.465	0.488	0.471	0.494	0.456	0.470	0.423	0.447
'	.	0.465	.	0.462	.	0.468	.	0.459	.	0.423
252	0.440	0.440	0.439	0.477	0.445	0.443	0.447	0.445	0.400	0.401
307	0.440	0.440	0.439	0.437	0.445	0.443	0.433	0.431	0.376	0.377
357	0.413	0.411	0.411	0.409	0.418	0.416	0.420	0.419	0.354	0.354
406	0.413	0.411	0.411	0.409	0.418	0.416	0.411	0.410	0.325	0.326
446	0.413	0.411	0.411	0.409	0.418	0.416	0.404	0.402	0.299	0.030
476	0.413	0.411	0.411	0.409	0.418	0.416	0.398	0.397	0.275	0.275
484	0.413	0.411	0.411	0.409	0.418	0.416	0.397	0.396	0.253	0.253
524	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
748	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
933	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1037	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1078	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1183	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1290	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1345	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
1560	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
2114	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233
2144	0.382	0.381	0.374	0.372	0.382	0.380	0.390	0.389	0.232	0.233

Tabela 4.7 Estimativas dos dados não perturbados comparados com os dados perturbados no  $t$  igual a 63, para a perturbação do Tipo II, para os dados reais.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
2	0.976	0.976	0.976	0.976	0.977	0.977	0.977	0.977	0.977	0.970
4	0.953	0.953	0.953	0.953	0.954	0.954	0.954	0.950	0.954	0.954
28	0.930	0.930	0.930	0.930	0.931	0.931	0.931	0.931	0.931	0.931
30	0.907	0.907	0.907	0.907	0.908	0.908	0.908	0.908	0.908	0.908
32	0.883	0.883	0.883	0.883	0.885	0.885	0.885	0.885	0.885	0.885
36	0.860	0.860	0.860	0.860	0.862	0.862	0.862	0.862	0.862	0.862
41	0.837	0.837	0.837	0.837	0.839	0.839	0.839	0.839	0.839	0.839
42	0.814	0.814	0.814	0.814	0.816	0.816	0.816	0.816	0.816	0.816
49	0.790	0.790	0.790	0.790	0.793	0.793	0.793	0.793	0.793	0.793
52	0.767	0.767	0.767	0.767	0.770	0.770	0.770	0.770	0.770	0.770
53	0.744	0.744	0.744	0.744	0.747	0.747	0.747	0.747	0.747	0.747
57	0.720	0.720	0.720	0.720	0.724	0.724	0.724	0.724	0.724	0.724
62	0.697	0.697	0.697	0.697	0.701	0.701	0.701	0.701	0.701	0.701
63	0.674	.	0.674	.	0.678	.	0.678	.	0.678	.
63 <sup>c</sup>	.	0.697	.	0.697	.	0.701	.	0.698	.	0.677
72	0.651	0.674	0.651	0.673	0.655	0.677	0.655	0.698	0.655	0.654
77	0.627	0.650	0.627	0.649	0.632	0.653	0.632	0.653	0.632	0.631
79	0.604	0.625	0.604	0.625	0.609	0.629	0.609	0.629	0.609	0.608
81	0.581	0.601	0.581	0.601	0.586	0.606	0.586	0.606	0.586	0.585
81	0.558	0.577	0.551	0.577	0.563	0.582	0.563	0.582	0.563	0.562
84	0.534	0.553	0.534	0.553	0.540	0.558	0.540	0.558	0.540	0.054
108	0.511	0.529	0.511	0.529	0.517	0.534	0.517	0.534	0.517	0.517
132	0.488	0.505	0.488	0.505	0.494	0.511	0.494	0.511	0.494	0.494
140	0.465	0.481	0.465	0.481	0.471	0.487	0.471	0.487	0.471	0.471

Tabela 4.7 (continuação) Estimativas dos dados não perturbados comparados com os dados perturbados no  $t$  igual a 63, para a perturbação do Tipo II, para os dados reais.

$t$	$ea$	$ea'$	$km$	$km'$	$en$	$en'$	$ek$	$ek'$	$mk$	$mk'$
180	0.465	0.481	0.465	0.481	0.471	0.487	0.462	0.478	0.446	0.446
210	0.465	0.481	0.465	0.481	0.471	0.487	0.456	0.471	0.423	0.422
252	0.440	0.456	0.439	0.454	0.445	0.460	0.447	0.462	0.400	0.400
307	0.440	0.456	0.439	0.454	0.445	0.460	0.433	0.447	0.376	0.376
357	0.413	0.428	0.411	0.426	0.418	0.433	0.420	0.435	0.354	0.354
406	0.413	0.428	0.411	0.426	0.418	0.433	0.4117	0.4257	0.325	0.325
446	0.413	0.428	0.411	0.426	0.418	0.433	0.404	0.418	0.299	0.299
476	0.413	0.428	0.411	0.426	0.418	0.433	0.398	0.412	0.275	0.275
484	0.413	0.428	0.411	0.426	0.418	0.433	0.397	0.411	0.253	0.253
524	0.382	0.395	0.374	0.387	0.382	0.385	0.390	0.403	0.232	0.232
748	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232
933	0.382	0.395	0.374	0.387	0.382	0.385	0.390	0.403	0.232	0.232
1037	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232
1078	0.382	0.395	0.374	0.387	0.382	0.385	0.390	0.403	0.232	0.232
1183	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232
1290	0.382	0.395	0.374	0.387	0.382	0.385	0.390	0.403	0.232	0.232
1345	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232
1560	0.382	0.395	0.374	0.387	0.382	0.385	0.390	0.4037	0.232	0.232
2114	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232
2144	0.382	0.395	0.374	0.387	0.382	0.395	0.390	0.403	0.232	0.232

No Capítulo 5 apresentamos um estimador exponencial por partes de Kitchin que estamos propondo nesta dissertação, denominado de estimador modificado de Kitchin ( $mk$ ). Comparamos suas estimativas com os demais estimadores apresentados no Capítulo 3, e a situação delas em relação aos dados de transplantes renais.

## Capítulo 5

# Proposta de Um Novo Estimador Modificado do Tipo Exponencial Por Partes de Kitchin

Neste Capítulo estamos propondo um novo estimador modificado do tipo exponencial por partes de Kitchin. Apresentamos o desenvolvimento desta proposta, ressaltando as modificações implementadas neste estimador, em relação ao estimador de Kitchin.

Em Benze (1997), a taxa instantânea de falha de Kitchin, em um subintervalo  $I_m$  entre dois pontos de falhas consecutivas, é reescrita na forma:

$$\lambda e_m^{ek}(te_m) = \left[ \frac{1}{V(te_m^-) + \sum_{c=1}^{W_m} f_c(I_m)} \right] \frac{1}{\Delta te_m}, \quad (5.1)$$

onde

$f_c(I_m)$  = fração do comprimento do subintervalo  $I_m = (te_{m-1} - te_m]$  que corresponde à distância da  $c$ -ésima censura neste subintervalo, em relação ao ponto  $te_{m-1}$ . Devemos ressaltar que este fator é uma espécie de correção para as censuras.

Na composição desta taxa, observamos o fator

$$f_m = \sum_{c=1}^{W_m} f_c(I_m), \quad (5.2)$$

representando um efeito das censuras observadas no subintervalo  $I_m$  e o fator

$$f_p = \frac{1}{\Delta te_m}, \tag{5.3}$$

correspondente a uma ponderação pelo inverso do tamanho do subintervalo

Na interpretação de Kitchin (1980), a taxa (5.1) é dada pelo inverso da área  $A_{ek}$ , sob a curva do processo de contagem de falhas e censuras, dentro dos limites  $te_{m-1}$  e  $te_m$  do subintervalo  $I_m$ . Isto é,

$$\lambda e_m^{ek}(te_m) = \frac{1}{A_{ek}}, \tag{5.4}$$

onde

$$A_{ek} = V(te_{m-1})[tc_1 - te_{m-1}] + V(tc_1)[tc_2 - tc_1] + \dots + V(tc_{w_m})[tc_{w_m} - tc_{w_m-1}], \tag{5.5}$$

conforme é ilustrado genericamente na Figura 5.1, a seguir.

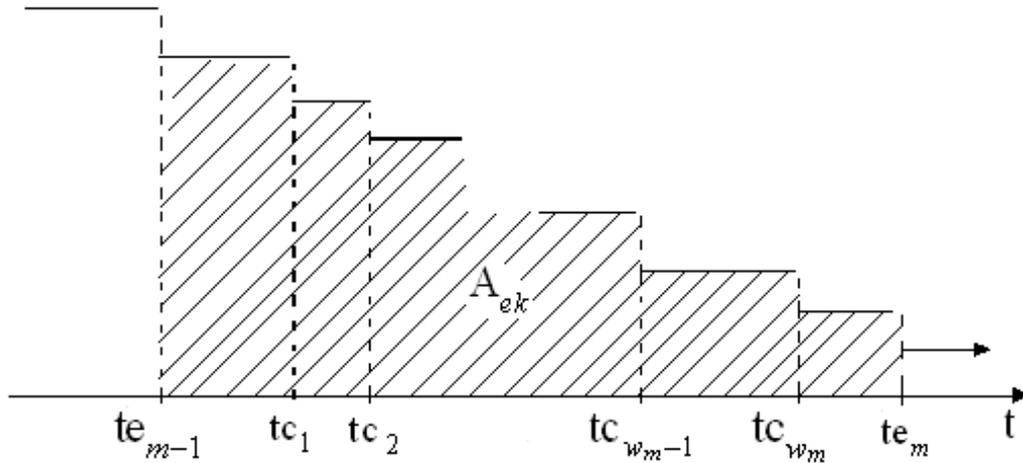


Figura 5.1 Representação gráfica da área sob a curva de uma realização hipotética do processo de contagem de falhas ( $m$ ) e censuras ( $c$ ), em um subintervalo  $I_m$ .

Na procura de um estimador alternativo da taxa instantânea de falhas no subintervalo  $I_m$  e com a finalidade de expandir para os demais estimadores a interpretação geométrica evidenciada em (5.4), Benze (1997) definiu as áreas:

$$A_{en} = V(te_m^-)\Delta te_m, \tag{5.6}$$

onde

$A_{en}$  representa a área sob a curva do processo de contagem, referente ao estimador de Nelson e

$$A_{ea} = V(te_{m-1})\Delta te_m, \tag{5.7}$$

onde

$A_{ea}$  representa a área sob a curva do processo de contagem, referente ao estimador atuarial adaptado para pequenas amostras.

A idéia relacionada ao procedimento adotado por Benze (1997) foi determinar um estimador mediante a combinação destas interpretações geométricas. Uma saída natural foi então definir:

$$\bar{A} = \frac{A_{en} + A_{ea}}{2}. \tag{5.8}$$

A expressão (5.8) pode ser representada na forma de processos de contagem por:

$$\begin{aligned} \bar{A} &= \frac{A_{en} + A_{ea}}{2} \\ &= \left[ \frac{V(te_m^-) + V(te_{m-1})}{2} \right] \Delta te_m \\ &= \left[ \frac{V(te_{m-1}) - W(I_m) + V(te_{m-1})}{2} \right] \Delta te_m \\ &= \left[ \frac{2V(te_{m-1})}{2} - \frac{W(I_m)}{2} \right] \Delta te_m \\ &= \left[ V(te_{m-1}) - \frac{W(I_m)}{2} \right] \Delta te_m. \end{aligned}$$

ou

$$\begin{aligned}
 \bar{A} &= \frac{A_{en} + A_{ea}}{2} \\
 &= \left[ \frac{V(te_m^-) + V(te_{m-1})}{2} \right] \Delta te_m \\
 &= \left[ \frac{V(te_m^-) + V(te_m^-) + W(I_m)}{2} \right] \Delta te_m \\
 &= \left[ \frac{2V(te_m^-)}{2} + \frac{W(I_m)}{2} \right] \Delta te_m \\
 &= \left[ V(te_{m-1}) + \frac{W(I_m)}{2} \right] \Delta te_m.
 \end{aligned}$$

Usando estas áreas, Benze definiu as taxas instantâneas de falhas correspondentes aos estimadores atuarial, de Kaplan-Meier e de Nelson, equivalentes à (5.4) e, por comparação analítica de todas essas taxas, propôs o estimador:

$$\lambda_e^{mk}(te_m) = \frac{\frac{1}{A_{en}} + \frac{1}{\bar{A}}}{2}. \tag{5.9}$$

Substituindo (5.6) e (5.8) em (5.9), obtém-se:

$$\lambda_e^{mk}(te_m) = \frac{1}{2} \left[ \frac{1}{V(te_m^-)} + \frac{1}{V(te_m^-) + \frac{W_m}{2}} \right] \frac{1}{\Delta te_m}. \tag{5.10}$$

onde

$$f_m^* = \frac{1}{V(te_m^-) + \frac{W_m}{2}}, \tag{5.11}$$

pode ser interpretado como um fator, já observado na forma funcional do estimador atuarial, representando o efeito conjunto das censuras (indicadas por  $\frac{W_m}{2}$ ) e da posição do subintervalo no seguimento do processo (indicada por  $V(te_m^-)$ ). Assim, é convencionalizado que metade das censuras "observadas" em um dado subintervalo só venham a abandonar efetivamente o estudo em tempos posteriores ao do limite superior  $te_m$  deste subintervalo.

Nestas condições, podemos reescrever as taxas (5.1) e (5.10) nas formas:

$$\lambda_e^{ek}(te_m) = \left( \frac{1}{V(te_m^-) + f_m} \right) f_p. \tag{5.12}$$

e

$$\lambda e_m^{mk}(te_m) = \frac{1}{2} \left( \frac{1}{V(te_m^-)} + f_m^* \right) f_p. \quad (5.13)$$

Os correspondentes estimadores exponencial por partes (PEXE) da curva de sobrevivência, em um tempo  $t$ ,  $te_{m-1} < t \leq te_m$ , são dados genericamente por:

$$Se^{xx}(t) = \exp \left\{ - \left[ (t - te_m) \lambda e_m^{xx}(te_m) + \sum_{i=1}^{m-1} \Delta te_i \lambda e_i^{xx}(te_i) \right] \right\}, \quad (5.14)$$

onde o sobrescrito  $xx$  será igual a  $ek$  ou  $mk$ , dependendo da taxa, (5.12) ou (5.13), respectivamente, que esteja sendo utilizada.

Neste trabalho, observamos nas simulações iniciais que as estimativas do tipo (5.14), usando a taxa (5.13), denotada pela taxa de Benze (1997), não apresentaram bom desempenho com relação ao vício e erro médio quadrático. Com a taxa (5.12), denotada pela taxa de Kitchin, esse desempenho foi melhor, inclusive apresentou melhor resistência à perturbação do Tipo I, verificados para os demais estimadores estudados no Capítulo 3.

Na taxa (5.12), quando é mantido o fator  $f_m$  e imaginadas duas ou mais situações, com a mesma quantidade  $W_m$  de censuras, no mesmo subintervalo  $I_m$ , o valor da taxa (5.12) será maior se as censuras estiverem localizadas mais para o início do subintervalo, quando comparado, por exemplo, com a situação destas censuras localizadas mais para o final do mesmo subintervalo.

Na hipótese de tendência da identificação mais tardia das censuras em cada subintervalo, isso acarretará superestimativas da curva de sobrevivência. Isto é, o decaimento da curva estimada será mais lento do que a realidade.

Este problema poderá se agravar com o aumento da porcentagem de censuras.

Uma observação mais cuidadosa das expressões (5.12) e (5.13) e suas componentes ( $f_m$  e  $f_m^*$ ) poderá conduzir a uma nova taxa de risco de falha, visando amenizar esse problema.

A simples substituição, em (5.12), do fator  $f_m$  pelo fator  $f_m^*$  é uma tentativa de amenizar este problema. Pois  $f_m^*$  manter-se á constante, independentemente das posições, dentro do subintervalo, julgadas como o momento de ocorrência das censuras. Além disso, ao adotarmos o fator  $f_m^*$ , levamos em conta, na estimativa da curva de sobrevivência no subintervalo, o efeito de somente metade das censuras, ao invés do total  $W_m$ . Com isso,

supomos que metade das observações classificadas como censuras no subintervalo sejam consideradas como potenciais tempos de falha neste subintervalo. Como consequência pode ocorrer uma diminuição no valor do denominador de (5.12) que leve ao aumento da taxa de falha e provoque maior decaimento da estimativa da curva de sobrevivência. Dando continuidade a esta análise, ainda, da substituição de (5.12), em (5.14), com a implementação da mudança de  $f_m$  por  $f_m^*$ , o primeiro termo dentro da expressão exponencial em (5.14) pode ser escrito na forma:

$$\begin{aligned} (t - te_m)\lambda e_m^*(te_m) &= (t - te_m) \left[ \frac{1}{V(te_m^-) + f_m^*} \right] f_p \\ &= \frac{t - te_m}{\Delta te_m} \left[ \frac{1}{V(te_m^-) + f_m^*} \right]. \end{aligned} \tag{5.15}$$

Para  $t < t \leq te_m$  tem-se  $0 < \frac{te_{m-1} - te_m}{\Delta te_m} \leq 1$ , o que prevê, na curva de sobrevivência estimada decaimentos mais lentos, quanto maior for o comprimento do subintervalo, contribuindo também para a possibilidade da produção de superestimativas. Se, em (5.15),  $f_p$  fosse substituído por  $\frac{1}{2}f_p$ , como é sugerido pela expressão (5.13), essa situação se agravaria.

Para aumentar a velocidade de decaimento na estimativa da curva de sobrevivência com o aumento do tamanho dos subintervalos  $I_m$  podemos aumentar a estimativa da taxa acumulada de falha, dada por

$$\Lambda^*(t) = (t - te_m)\lambda e_m^*(te_m) + \sum_{i=1}^{m-1} \Delta te_i \lambda e_i^*(te_i), \tag{5.16}$$

substituindo a ponderação  $\frac{t-te_m}{\Delta te_m}$  por  $\frac{t-te_m}{K}$ , onde  $K$  é uma constante menor do que  $\Delta te_m$ . Um critério de escolha de  $K$  é considerar a metade dos comprimentos  $t - te_m$ , isto é  $\frac{t-te_m}{2}$ . De certa forma, isto já parecia estar sendo sugerido pela expressão (5.13). Realmente, as simulações iniciais indicaram  $K = 2$  como sendo a melhor escolha dentre outros valores próximos, como por exemplo  $K = 1$  ou  $K = 3$ .

Nessas condições, propomos uma nova forma modificada do estimador da taxa instantânea de falhas de Kitchin, dada por:

$$\lambda e_m^{**}(te_m) = \frac{1}{2} \left[ \frac{1}{V(te_m^-) + \frac{1}{V(te_m^-) + \frac{W_m}{2}}} \right]. \quad (5.17)$$

a qual é denotada por ( $mk$ ).

Para uma interpretação análoga a dada pela expressão (5.9) podemos fazer:

$$\begin{aligned} \lambda e_m^{**}(te_m) &= \frac{1}{2} \left[ \frac{1}{\frac{V(te_m^-)\Delta te_m}{\Delta te_m} + \frac{\Delta te_m}{(V(te_m^-) + \frac{W_m}{2})\Delta te_m}} \right] \\ &= \frac{1}{2} \left[ \frac{1}{\frac{A_{en}}{\Delta te_m} + \frac{\Delta te_m}{A}} \right] \\ &= \frac{1}{\frac{A_{en} + \Delta te_m}{\Delta te_m} + \frac{\Delta te_m}{A}}. \end{aligned} \quad (5.18)$$

As modificações implementadas neste estimador, em relação ao estimador de Kitchin, parecem torná-lo menos vulnerável às superestimativas decorrentes, por exemplo, à perturbação do Tipo II, referente a realização da troca de uma falha por uma censura. Como podemos ver na seção 6.4, nos estudos de simulação apresentados, que esta robustez foi comprovada.

É importante ressaltar que as perturbações realizadas, Tipo I e II, produzem o aumento, tanto do tamanho do subintervalo, como do número de censuras nele contidas, gerando essas superestimativas. Além disso, essas modificações podem amenizar o efeito da ocorrência de grandes porcentagens de censuras comuns na maioria dos conjuntos de dados de sobrevivência.

Na Figura 5.2, apresentamos a estimativa da curva de sobrevivência segundo o novo estimador modificado do tipo exponencial por partes de Kitchin,  $mk$ , para os dados reais citados na seção 3.2.

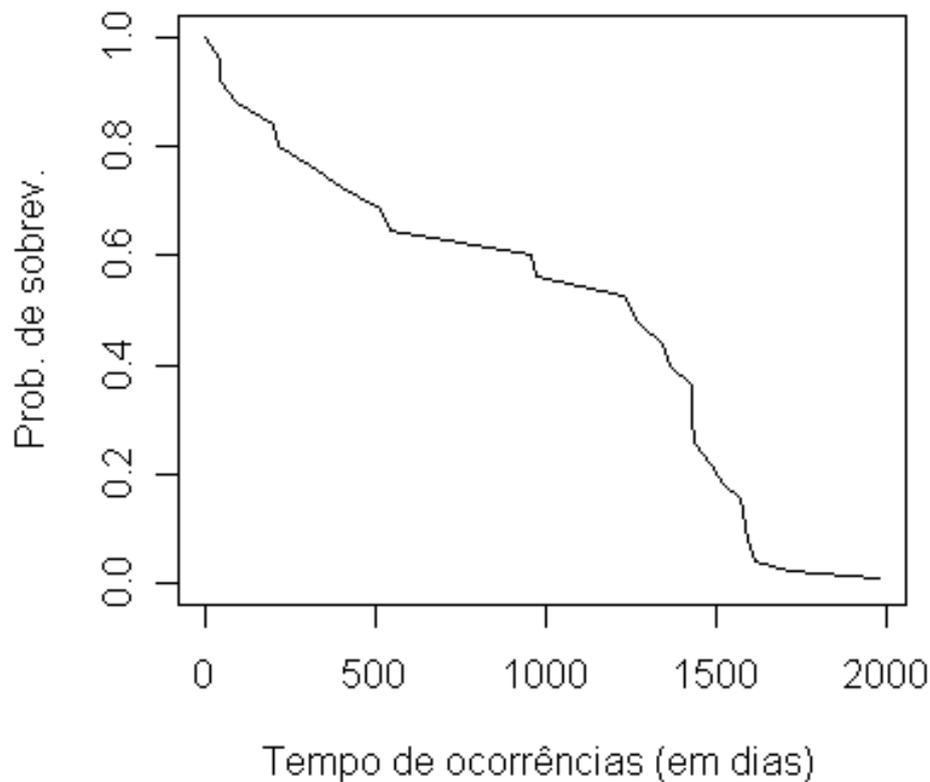


Figura 5.2 Estimativa da curva de sobrevivência  $mk$  para os dados de transplantes renais.

Observamos que a curva decai mais rapidamente e a sua cauda superior apresenta-se mais próxima do eixo das abscissas, quando comparada com a dos outros estimadores. A seguir, comparamos, usando o exemplo de rins o novo estimador com os anteriores. Ver a Figura 5.3, onde comparamos as estimativas de alguns dos estimadores apresentados na introdução deste trabalho, entre estes o estimador atuarial, Littell (1952), citado em Kaplan e Meier (1958), o estimador de Kaplan-Meier (1958), o estimador de Nelson (1970), o estimador exponencial por partes de Kitchin (1980) e o estimador proposto neste trabalho, modificado de Kitchin. Na Figura 5.4, são acrescentados os correspondentes limites de Peterson.

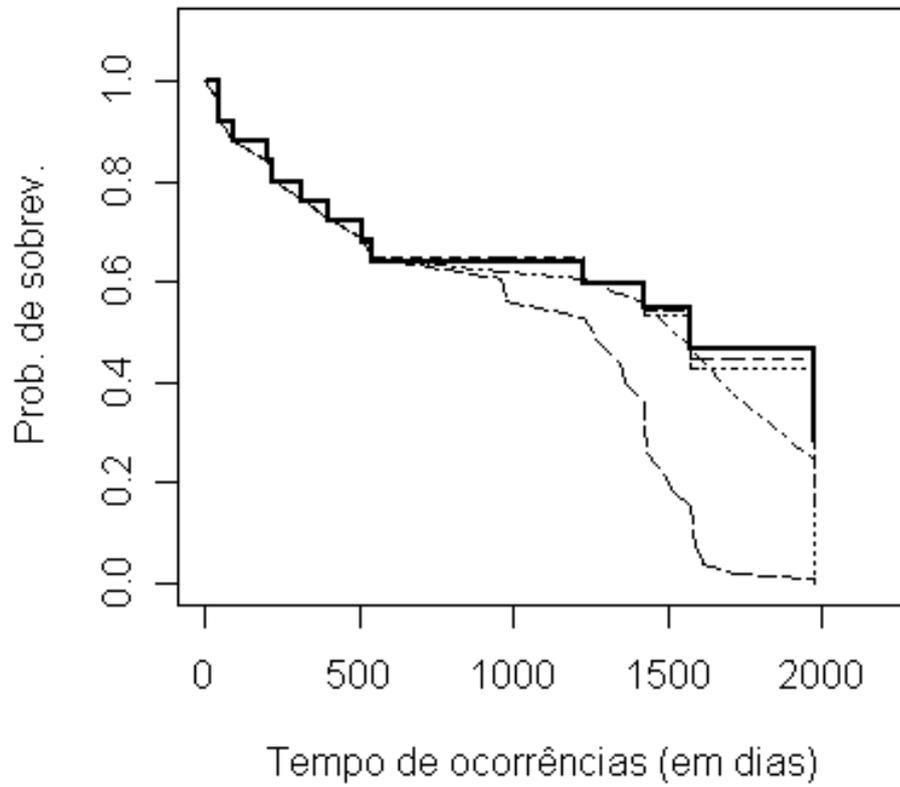


Figura 5.3 Comparação das estimativas  $ea$ ,  $km$ ,  $en$  e  $ek$ , e situação delas em relação aos dados de transplantes renais. As estimativas  $ea$  são apresentadas por (—), as  $en$  são representadas por (---), as  $km$  são representadas por (.....), as  $ek$  são representadas por (-.-.-) e as  $mk$  são representadas por (-.-.-).

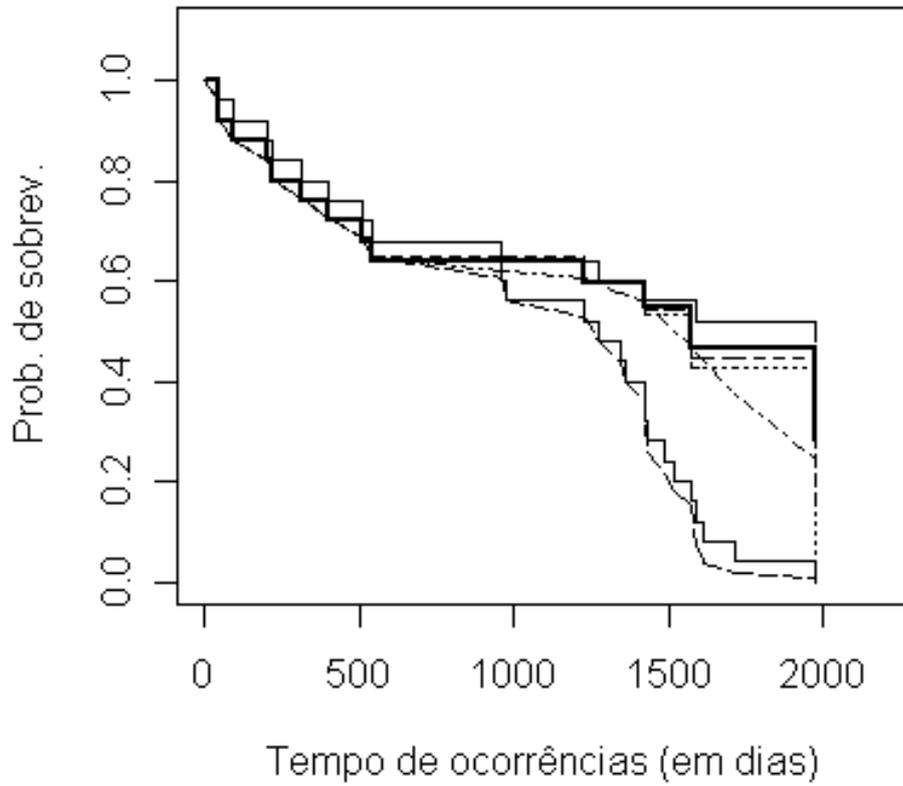


Figura 5.4 Comparação das estimativas  $ea$ ,  $km$ ,  $en$  e  $ek$ , e situação delas em relação aos dados de transplantes renais. As estimativas  $ea$  são representadas por (—), as  $en$  são representadas por (---), as  $km$  são representadas por (....), as  $ek$  são representadas por (-.-.-) e as  $mk$  são representadas por (- - - -) e os limites de Peterson são representados por (—).

Tabela 5.1 Estimativas  $ea$ ,  $km$ ,  $en$ ,  $ek$ ,  $mk$  e dos limites de Peterson.  $LI$  representando o limite inferior e  $LS$  o limite superior da curva de sobrevivência para os dados de transplantes renais.

$t$	$ea$	$km$	$en$	$ek$	$mk$	$LI$	$LS$
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.960	0.960	0.960	0.960	0.960	0.960	0.960
43	0.920	0.920	0.961	0.921	0.921	0.920	0.920
89	0.880	0.880	0.882	0.882	0.882	0.880	0.880
201	0.840	0.840	0.843	0.843	0.843	0.840	0.840
215	0.800	0.800	0.803	0.803	0.803	0.800	0.800
308	0.760	0.760	0.764	0.764	0.764	0.760	0.760
396	0.720	0.720	0.725	0.725	0.725	0.720	0.720
512	0.680	0.680	0.686	0.686	0.686	0.680	0.680
542	0.640	0.640	0.647	0.647	0.647	0.640	0.640
956 <sup>c</sup>	0.640	0.640	0.647	0.621	0.603	0.600	0.640
970 <sup>c</sup>	0.640	0.640	0.647	0.621	0.563	0.560	0.640
1227	0.597	0.594	0.602	0.606	0.563	0.520	0.600
1272 <sup>c</sup>	0.597	0.594	0.602	0.594	0.526	0.480	0.600
1343 <sup>c</sup>	0.597	0.594	0.602	0.575	0.479	0.440	0.600
1362 <sup>c</sup>	0.597	0.594	0.602	0.570	0.436	0.400	0.600
1424	0.545	0.534	0.545	0.555	0.397	0.360	0.560
1426 <sup>c</sup>	0.545	0.534	0.5452	0.554	0.361	0.320	0.560
1434 <sup>c</sup>	0.545	0.534	0.545	0.549	0.304	0.280	0.560
1483 <sup>c</sup>	0.545	0.534	0.545	0.520	0.256	0.240	0.560
1518 <sup>c</sup>	0.545	0.534	0.545	0.500	0.216	0.200	0.560
1570	0.467	0.427	0.446	0.471	0.182	0.160	0.520
1588 <sup>c</sup>	0.467	0.427	0.446	0.458	0.153	0.120	0.520
1609 <sup>c</sup>	0.467	0.427	0.446	0.442	0.076	0.080	0.520
1713 <sup>c</sup>	0.467	0.427	0.446	0.372	0.037	0.040	0.520
1973	0.280	0.000	0.164	0.241	0.018	0.000	0.480

Observamos que todas as estimativas estão dentro dos limites de Peterson, conforme já mencionamos na seção 3.10. O estimador  $mk$ , comparado com os demais, subestima as probabilidades de sobrevivência. Enquanto o  $ek$  apresenta-se na metade da região determinada pelos limites de Peterson, as estimativas  $mk$  apresentam-se próximas das estimativas do limite inferior  $\hat{L}I$ .

No próximo Capítulo apresentamos os procedimentos de reamostragem para verificar a precisão dos estimadores estudados. Citamos os procedimentos de geração de amostras e os resultados deste estudo de simulação.

# Capítulo 6

## Procedimentos de Reamostragem para Verificar a Precisão dos Estimadores

Neste Capítulo apresentamos os procedimentos de reamostragem para verificar a precisão dos estimadores estudados nesta dissertação. Citamos o procedimento que utilizamos para realizarmos as perturbações do Tipo I e do Tipo II, os procedimentos de geração de amostras e os resultados deste estudo de simulação.

Um dos principais objetivos da Estatística é quantificar o erro cometido ao estimar um determinado parâmetro  $\theta$  de interesse. Uma forma de obter medidas de incerteza que expressem este erro, é estimar o erro padrão da estimativa pontual,  $\hat{\theta}$ . Entretanto, essas medidas nem sempre são obtidas analiticamente, pois podem apresentar expressões altamente complexas. Por outro lado, a construção de intervalos de confiança baseado na teoria assintótica, exige que as amostras sejam suficientemente grandes, a fim de garantir a validade das aproximações em questão, caso contrário, a eficácia do método é seriamente comprometida.

O método bootstrap é uma alternativa eficiente, através da qual obtém-se a estimativa do erro padrão de  $\hat{\theta}$ , livre de complexidades algébricas e possibilita a obtenção de intervalos de confiança sem a necessidade de pressupostos sobre a distribuição do estimador. Proposto por Bradley Efron, (1979), o método bootstrap foi inicialmente utilizado para encontrar o erro padrão de um parâmetro a ser estimado. Classificado no grupo de

métodos computacionais intensivos, tem sido um instrumento alternativo eficiente em várias situações, principalmente quando as amostras são de tamanho pequeno ou moderado. O bootstrap é essencialmente um método de reamostragem não paramétrica que posteriormente foi desenvolvido na forma paramétrica. O bootstrap paramétrico reamostra a partir de um modelo probabilístico cujos parâmetros são estimados a partir da amostra original. O bootstrap não paramétrico faz a reamostragem com reposição, a partir da amostra aleatória simples original.

A estimação intervalar via método bootstrap é indicada quando há dificuldades analíticas de se encontrar intervalos de confiança exatos e, quando aproximados, existe a necessidade de ter grandes amostras a fim de garantir a validade das aproximações em questão.

Com a finalidade de verificarmos a precisão dos estimadores estudados, comparamos os seus desempenhos calculando, em média, o vício, o erro quadrático médio e a probabilidade de cobertura para os intervalos de confiança (IC) 95% bootstrap.

## 6.1 O estimador bootstrap

O método bootstrap tem por base a idéia de que o pesquisador pode tratar a sua amostra como se esta fosse a população que deu origem aos dados e usar amostragem com reposição de sua amostra original para gerar pseudo-amostras e, a partir destas, estimar características de interesse de certas estatísticas. Vários esquemas diferentes de bootstrap têm sido propostos e muitos deles apresentam bom desempenho em uma ampla variedade de situações. A técnica de reamostragem bootstrap, Efron & Tibshirani (1993) foi recentemente desenvolvida para realizar certas inferências estatísticas. Requer modernos computadores para simplificar os freqüentes emaranhados de cálculos da teoria, com o mínimo de suposições matemáticas. Uma amostra bootstrap é obtida reamostrando-se aleatoriamente  $n$  vezes com reposição as observações da amostra original obtendo-se  $l$  replicações. Esse método de simulação é muito utilizado para estudar tendências, permitindo fazer inferências para a população que originou os dados. Uma simulação não-paramétrica bootstrap assume que  $X_1, \dots, X_n$  são variáveis aleatórias independentes e identicamente distribuídas, (i.i.d.) e a reamostra é feita sobre  $X_1, \dots, X_n$ . Uma simulação paramétrica

bootstrap assume que  $X_1, \dots, X_n$  são variáveis aleatórias, (i.i.d.) com distribuição  $F$  e a reamostra  $X_1, \dots, X_n$  é feita diretamente de  $F$ .

Quando se trabalha com a teoria assintótica, algumas questões surgem naturalmente. Será que o tamanho da amostra é suficientemente grande para garantir uma boa qualidade da estimação? Ou, ao considerar medidas de dependência entre duas variáveis, qual a variabilidade de suas estimativas? Como tal variabilidade mudaria as conclusões feitas? Pode-se responder as questões acima obtendo-se uma variabilidade empírica das estimativas por meio de esquemas de bootstrap de reamostragem.

### 6.1.1 O estimador bootstrap do erro padrão

Seja uma amostra aleatória baseada em  $n$  observações independentes  $x_1, x_2, \dots, x_n$ . O erro padrão de uma média baseada nesta amostra é estimada pela expressão:

$$\hat{e}p(\bar{x}) = \sqrt{\frac{s^2}{n}}, \quad (6.1)$$

onde

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (6.2)$$

é o estimador não viciado da variância. O erro padrão não é uma estimativa de uma quantidade pertinente de uma população, mas uma medida de incerteza da média amostral vista como uma estimativa da média populacional (1991). Pela expressão (6.1), nota-se que a magnitude desta incerteza diminui conforme o tamanho da amostra ( $n$ ) aumenta.

Pode-se generalizar este conceito da seguinte maneira: Seja uma variável aleatória  $X$  com distribuição  $F$ , sendo sua esperança denotada por  $\mu_F$  e sua variância denotada por  $\sigma_F^2$ . Sendo  $x$  um valor pertencente a esta variável aleatória e usando-se a notação  $x \sim (\mu_F, \sigma_F^2)$ , pode-se escrever  $F \longrightarrow (x_1, x_2, \dots, x_n)$  para indicar que  $x = (x_1, x_2, \dots, x_n)$  é uma amostra aleatória de tamanho  $n$  obtida de uma população com distribuição  $F$ . A média desta amostra é um valor da variável aleatória, com distribuição  $F'$ , cuja esperança é  $\mu_{F'} = \mu_F$  e variância  $\sigma_{F'}^2 = \frac{\sigma_F^2}{n}$ , ou seja,  $\bar{X} \sim (\mu_F, \frac{\sigma_F^2}{n})$ . Portanto,  $X$  e  $\bar{X}$  tem a mesma esperança; entretanto, a variância de  $\bar{X}$  é  $n^{-1}$  vezes a variância de  $X$ , isto é:

$$ep_{F'}(\bar{X}) = \sqrt{\text{var}_{F'}(\bar{X})} = \frac{\sigma_F}{\sqrt{n}}. \quad (6.3)$$

A expressão (6.1) fornece um estimador para (6.3). Entretanto, nem todos os estimadores têm expressões de fácil manejo para seu respectivo erro padrão, como (6.1). Isto significa que o trabalho de encontrar medidas de precisão para outros estimadores, que não a média, pode ser algo bastante complicado.

Sob certas condições, o teorema central do limite diz que a distribuição dos valores  $\bar{x}$  de  $\bar{X}$  é aproximadamente normal quando  $n$  é grande, ou seja:

$$\bar{X} \sim N\left(\mu_F, \frac{\sigma_F^2}{n}\right), \quad (6.4)$$

e, conseqüentemente,

$$\frac{\bar{X} - \mu_F}{\frac{\sigma_F}{\sqrt{n}}} \sim N(0, 1). \quad (6.5)$$

Usando uma tabela de distribuição normal padrão, de (6.5) pode-se escrever, por exemplo, a probabilidade

$$P\left(\left|\bar{X} - \mu_F\right| < \frac{2\sigma_F}{\sqrt{n}}\right) = 0.954, \quad (6.6)$$

onde o erro padrão mostra-se uma medida útil de precisão da estimativa da média. A probabilidade da distância entre a média amostral e a populacional ser de pelo menos dois erros padrão é de 0.954. Verifica-se assim que, quanto menor o erro padrão da média, a expressão (6.6) melhor sugere uma proximidade entre as médias amostral e populacional.

Uma vantagem do bootstrap é que esta técnica não depende inteiramente do teorema central do limite, já que, em suas aplicações, as medidas de precisão são obtidas diretamente dos dados, Efron & Tibshirani (1993).

Observada uma amostra aleatória de tamanho  $n$ , com distribuição  $F$ , define-se uma função distribuição empírica  $\hat{F}_n$  como uma distribuição discreta, que atribui probabilidade  $n^{-1}$  a cada  $x_i$ ,  $i = 1, 2, \dots, n$ . Uma amostra bootstrap  $x^* = (x_1^*, x_2^*, \dots, x_n^*)$  é obtida reamostrando-se aleatoriamente  $n$  vezes, com reposição, as observações  $x = (x_1, x_2, \dots, x_n)$ , de tal forma que se verifica  $\hat{F}_n \rightarrow (x_1^*, x_2^*, \dots, x_n^*)$ .

O estimador bootstrap do  $ep_F(\hat{\theta})$  – erro padrão de  $\hat{\theta}$  – é um estimador que utiliza a função distribuição empírica de  $\hat{F}$ , ao invés da função desconhecida  $F$ . Ou seja, o estimador bootstrap de  $ep_F(\hat{\theta})$  é o erro padrão de  $\hat{\theta}$  para conjuntos de dados de tamanho  $n$  amostrados aleatoriamente de  $\hat{F}$ . Seleccionadas  $L$  amostras bootstrap,  $x^{*1}, x^{*2}, \dots, x^{*L}$ , de forma independente, estima-se em cada uma delas  $\hat{\theta}^{*l} = s(x^{*l})$ , onde obtemos o desvio padrão calculado na amostra  $x^{*l}$ , com  $l = 1, 2, \dots, L$ . Uma expressão para o estimador bootstrap do erro padrão da estatística  $\hat{\theta}$  é dada por:

$$\hat{ep}_{boot} = \sqrt{\frac{\sum_{l=1}^L (s(x^{*l}) - s(\cdot))^2}{L - 1}}, \quad (6.7)$$

sendo que

$$s(\cdot) = \frac{\sum_{l=1}^L (s(x^{*l}))}{L}. \quad (6.8)$$

Chama-se de estimador bootstrap ideal de  $ep_F(\hat{\theta})$  o limite de  $\hat{ep}_{boot}$  quando  $L$  tende para o infinito, ou seja,

$$\lim_{L \rightarrow \infty} \hat{ep}_{boot} = ep_{\hat{F}}(\hat{\theta}^*). \quad (6.9)$$

O estimador bootstrap ideal e sua aproximação, dada pela expressão (6.7), são chamados estimadores bootstrap não paramétricos, já que se baseiam em  $\hat{F}$ , um estimador não paramétrico de  $F$ . Um estimador bootstrap paramétrico do erro padrão é baseado num estimador  $\hat{F}$  de  $F$  derivado de um modelo paramétrico. Por exemplo, ao invés de se estimar  $F$  pela função distribuição empírica  $\hat{F}$ , pode-se assumir que a população tenha uma distribuição conhecida.

## 6.2 Estimação intervalar

### 6.2.1 Estimação por intervalo via bootstrap

Uma simples estimativa pontual de um parâmetro de interesse pode não ser suficiente para que sejam obtidas informações pertinentes sobre um determinado estudo. São necessárias medidas de precisão desta estimativa, para que as conclusões possam ser mais confiáveis. O erro padrão e os intervalos de confiança são medidas de precisão que fornecem uma avaliação do erro que se comete na estimação pontual.

### 6.2.2 Intervalos de confiança

Sejam  $x = (x_1, x_2, \dots, x_n)$  uma amostra aleatória composta por  $n$  observações com uma distribuição  $F$ , isto é  $x$  é uma observação do vetor aleatório  $X = (X_1, X_2, \dots, X_n)$ , onde cada componente  $X_i$  é uma variável aleatória com distribuição  $F$ . Os intervalos de confiança são escritos na forma:

$$(L_I(x); L_S(x)), \quad (6.10)$$

onde  $(L_I(x))$  e  $(L_S(x))$  são os limites inferior e superior do intervalo.

Tem-se que

$$P_F[L_I(x) \leq \theta \leq L_S(x)] = 1 - 2\alpha, \quad (6.11)$$

ou seja, de cada 100 possíveis realizações de um experimento com o objetivo de gerar uma estimativa  $\hat{\theta}$  de  $\theta$ , com todas as amostras de tamanho  $n$ ,  $(1 - 2\alpha)100\%$  produzirão intervalos de confiança que deve conter  $\theta$ , o verdadeiro parâmetro, com a probabilidade  $(1 - 2\alpha)$  especificada.

A amostra obtida em qualquer experimento é uma entre várias amostras que poderiam ser geradas pelo mesmo processo de amostragem. Portanto, qual a confiança da estimativa  $\hat{\theta}$  de  $\theta$ ?

A quantidade  $(1 - 2\alpha)$  é denominada coeficiente de confiança, e auxilia o pesquisador a responder a esta questão. O lado esquerdo de (6.11) é dito probabilidade de cobertura, sendo que comumente  $0 \leq \alpha \leq 0.05$ .

Geralmente, intervalos de confiança são construídos com caudas iguais, isto é, satisfazem as expressões

$$P_F[\theta \leq L_I(X)] = \alpha \quad \text{e} \quad P_F[\theta \geq L_S(X)] = \alpha. \quad (6.12)$$

É importante ressaltarmos que (6.12) implica em (6.11), mas nem sempre (6.11) implica em (6.12).

Intervalos de confiança como o descrito em (6.11) são chamados exatos, quando a probabilidade em questão é exatamente igual a  $(1-2\alpha)$ . Entretanto, nem sempre é possível construir intervalos com tal propriedade. Os intervalos de confiança mais utilizados são os chamados aproximados, a saber

$$P_F[L_I(x) \leq \theta \leq L_S(x)] \approx 1 - 2\alpha, \quad (6.13)$$

sendo,

$$L_I = \bar{S}e - p(0.95) \hat{e}p_{boot}(Se^{xx}(d_j)), \quad (6.14)$$

e

$$L_S = \bar{S}e + p(0.95) \hat{e}p_{boot}(Se^{xx}(d_j)), \quad (6.15)$$

onde  $(^{xx})$  correspondem às duas letras da sigla do estimador a qual estejamos nos referindo: *ea* para o estimador atuarial; *km* para o estimador de Kaplan-Meier; *en* para o estimador de Nelson-Aalen, *ek* para o estimador de Kitchin e *mk* para o estimador proposto modificado de Kitchin. Temos a estimativa do erro padrão das estimativas nos decis da função de sobrevivência segundo o estimador *xx* dada por

$$\hat{e}p_{boot}(Se^{xx}(d_j)) = \sqrt{\frac{\sum_{l=1}^L (Se_l^{xx}(d_j) - \bar{S}e^{xx}(.))^2}{L-1}}, \quad (6.16)$$

sendo  $p(0.95)$  o 95<sup>o</sup>-ésimo percentil obtido das  $Se_l^{xx}(d_j)$  ordenadas em ordem crescente,  $d_j$  é o  $j$ -ésimo decil determinado na curva de sobrevivência teórica considerada e  $\bar{S}e^{xx}(. ) = \frac{\sum_{l=1}^L (Se_l^{xx}(d_j))}{L}$ .

Para se construir intervalos como os descritos acima, utiliza-se a teoria assintótica.

Sua limitação principal baseia-se no fato de se necessitar de uma amostra grande o suficiente para garantir a validade das aproximações em questão; caso contrário, a eficácia do método compromete-se demasiadamente. Intervalos de confiança exatos muitas vezes são construídos através de soluções analíticas nem sempre simples, enquanto os intervalos aproximados dependem de aproximações assintóticas que podem não ser alcançadas.

Os intervalos de confiança bootstrap são obtidos das estimativas obtidas em cada decil nas  $l$  reamostras geradas, ou seja,  $(s(x^{*1}), (s(x^{*2}), \dots, (s(x^{*l}))$  onde  $l = 1, \dots, L$ . Assim, em média  $(s(x^{*l}) = \frac{\sum_{l=1}^L (s(x^{*l}))}{L} = (s(x^{*l}) \pm (p_{95}) * \hat{e}p_{boot}$ .

### 6.3 Procedimento de geração das amostras

Para compararmos o desempenho dos estimadores apresentados, utilizamos um procedimento de simulação usando o método bootstrap paramétrico, onde reamostramos  $L$  vezes, uma amostra de tamanho  $n$ , de uma distribuição teórica conhecida (por exemplo, distribuição exponencial, distribuição Weibull,...) variando os parâmetros. Para isso, implementamos um algoritmo para gerar as amostras a partir dessas distribuições conhecidas e também para executar as etapas exigidas nos procedimentos de cálculo e comparação das estimativas dos estimadores. Para esta implementação utilizamos o software R (2005).

Em cada uma das  $l = 1, 2, \dots, L$  reamostras de tempos de sobrevivência, calculamos as estimativas das curvas de sobrevivência, obtidas por cada um dos estimadores ( $ea, km, en, ek$  e  $mk$ ). Dessa forma, ao final das  $L$  reamostras, obtivemos  $L$  vetores de tamanho  $n$ , correspondentes aos tempos de falhas e censuras com seus respectivos vetores, de tamanho  $n$  e as respectivas estimativas de sobrevivência.

Para podermos comparar os estimadores, calculamos os valores da função de sobrevivência teórica nos pontos correspondentes aos seus decis e depois estimamos a função de sobrevivência teórica em cada ponto dos decis, para cada reamostra e para cada estimador.

Para os estimadores que não admitem empates ( $ek$  e  $mk$ ) implementamos um mecanismo de comparação dos valores, durante a própria geração dos tempos, permitindo a identificação da ocorrência e o desempate dos valores utilizando acréscimos de tempos gerados de uma distribuição uniforme definida no intervalo  $(1/2, 1)$ . Este procedimento

assegura para cada amostra, bootstrap o estabelecimento exato da quantidade fixada de censuras.

Para cada um dos estimadores  $Se^{xx}(\cdot)$ , determinamos, respectivamente, em média, o vício, o erro médio quadrático e a probabilidade de cobertura para o intervalo de confiança 95%, em um ponto fixado  $x_i$ . Assim, levando-se em conta todas as  $N$  amostras na configuração, temos,

$$\text{Vício}^{xx} = \frac{1}{N} \sum_{j=1}^N [Se^{xx}(d_j) - Se(d_j)]_j, \quad (6.17)$$

onde  $Se^{xx}(d_j)$  são as estimativas nos decís da função de sobrevivência segundo o estimador  $xx$  e  $Se(d_j)$  são as verdadeiras probabilidades de sobrevivência ( $Se(d_j) = P(Te > d_j)$ ).

e

$$\text{EQM} = \{\text{var}[Se^{xx}(x_i)] + \text{Vício}^{xx}\}^2. \quad (6.18)$$

Para realizarmos as perturbações do Tipo I e do Tipo II, no procedimento de geração das amostras, além do descrito acima, implementamos no algoritmo os mecanismos para o efeito destas perturbações, os quais descreveremos a seguir.

Para a perturbação do Tipo I, por exemplo, implementamos o seguinte procedimento no algoritmo:

Passo 1: Temos um vetor de falhas e de censuras;

Passo 2: Sorteamos um tempo em que ocorreu falha, que esteja entre os primeiros 50% dos tempos gerados em que ocorreram as falhas e/ou censuras;

Passo 3: Sorteado o tempo de falha, alteramos a falha que ocorreu neste tempo para ocorrer no tempo após a primeira censura. Isto é, por exemplo, se o tempo de falha sorteado no passo 2 foi o 4, então a falha ocorrida neste tempo é indicada para ocorrer no tempo 7, se a primeira censura ocorreu no tempo 6;

Passo 4: O vetor recebe a falha alterada e continua o processo de execução das etapas exigidas nos procedimentos de cálculo e comparação das estimativas, de acordo com a configuração escolhida;

Restrição: Para a perturbação do Tipo I, se a primeira censura ocorrer no último tempo de observação, a perturbação do Tipo I não será realizada, devido a exigência

descrita no passo 3.

Para a perturbação do Tipo II, implementamos o seguinte procedimento no algoritmo:

Passo 1: Temos um vetor de falhas e de censuras;

Passo 2: Sorteamos um tempo em que ocorreu falha, que esteja entre os primeiros 50% dos tempos gerados em que ocorreram falhas e/ou censuras;

Passo 3: Sorteado o tempo de falha, alteramos a falha que ocorreu neste tempo por uma censura. Isto é, por exemplo, se o tempo de falha sorteado no passo 2 foi o 6, então a falha ocorrida nesse tempo será indicada para ocorrer uma censura, ou seja, no tempo 6 agora temos uma censura e não mais uma falha;

Passo 4: O vetor recebe a falha alterada por uma censura e continua o processo de execução das etapas exigidas nos procedimentos de cálculo e comparação das estimativas, de acordo com a configuração escolhida.

## 6.4 Resultados do estudo de simulação

Os resultados do estudo de simulação são condensados em Tabelas e Gráficos que são alocados no Apêndice A e B para mais clareza e fluidez da leitura.

Os valores registrados em Tabelas, em cada uma das seções de A.1 a A.13, no Apêndice A, foram calculadas utilizando configurações com  $L = 1000$  reamostras, de tamanhos  $n = 20$ , ou 50 ou 100, de dados simulados segundo o método descrito nas seções anteriores, com distribuições teóricas pré-determinadas, com funções de sobrevivência  $S_e$  e  $S_c$  [(Exponencial e Exponencial), (Exponencial e Weibull), e (Weibull e Weibull)] conforme valores escolhidos dos parâmetros e com as seguintes proporções de censuras  $r = 0.10$  ou 0.30 ou 0.50. No apêndice A, apresentamos o desempenho dos estimadores para situações específicas em que os tempos de falhas são supostos exponencialmente distribuídos e, depois são supostos com distribuição exponencial e/ou Weibull. Para cada uma dessas configurações, são determinados, em média, os resultados do vício e erro quadrático médio e os intervalos de confiança bootstrap das estimativas de sobrevivência nos decís da distribuição teórica, bem como os correspondentes resultados das probabilidades de cobertura.

Os registros em Tabelas, em cada uma das seções de B.1 a B.5 no Apêndice B, foram calculados utilizando configurações com  $L = 1000$  reamostras,  $n = 20$ , ou  $50$ , de dados simulados, com distribuições da família exponencial, conforme valores escolhidos dos parâmetros e com as seguintes proporções de censuras  $r = 0.10$ ,  $0.30$  ou  $0.50$ . Neste contexto, apresentamos, em média, os resultados do desempenho dos estimadores com base nos dados simulados na situação normal, isto é, não perturbados, comparados com os dados sujeitos às perturbações do Tipo I e do Tipo II. A seguir, apresentamos, em média, os resultados com relação ao vício e erro quadrático médio, os intervalos de confiança bootstrap, das estimativas de sobrevivência nos decís da distribuição teórica. Também apresentamos os resultados das probabilidades de cobertura e as diferenças relativas das estimativas, nas configurações especificadas, para as perturbações Tipo I e II, em relação à situação normal. São representados graficamente, via desenhos esquemáticos, as distribuições das diferenças relativas das estimativas, nas configurações consideradas das situações de perturbações em relação à situação normal. Em cada caso apresentamos as conclusões do resultado.

Para as situações específicas em que os tempos de falhas são supostos exponencialmente distribuídos, na ausência das perturbações e proporções de censuras  $0.10$ ,  $0.30$  e  $0.50$ , os estudos da simulação bootstrap indicaram, em média, um melhor desempenho, em relação ao vício e erro médio quadrático do estimador proposto, modificado de Kitchin ( $mk$ ) em relação aos demais. Os intervalos de confiança bootstrap, das estimativas de sobrevivência, determinadas para os decís da distribuição teórica, apresentaram probabilidades de cobertura que confirmaram estes resultados. Veja os resultados nas Tabelas de B 1.1 a B 7.3, nas configurações:  $(n = 20; exp(1/500); r = 0.10)$ ;  $(n = 50; exp(1/500); r = 0.10)$ ;  $(n = 50; exp(1/500); r = 0.30)$ ;  $(n = 100; exp(1/500); r = 0.10)$ ;  $(n = 50; exp(1/200); r = 0.30)$ ;  $(n = 20; exp(1/40); r = 0.30)$  e  $(n = 50; exp(1/40); r = 0.10)$ .

Segundo os métodos descritos neste Capítulo, com as distribuições  $Fe$  e  $Fc$  pertencentes às famílias exponencial e Weibull, na ausência das perturbações, e em amostras com proporções de censuras  $0.10$ ,  $0.30$  e  $0.50$ , o desempenho dos estimadores com dados simulados nas configurações  $(n = 20; exp(1/40); exp(1/500); r = 0.10)$  e  $(n = 50; exp(1/40); exp(1/500); r = 0.30)$ , (veja Tabelas de B 8.1 a B 9.3), indicaram em média melhor desempenho, do estimador proposto modificado de Kitchin ( $mk$ ) em relação ao

vício e erro médio quadrático.

Para as configurações, envolvendo a distribuição exponencial para os tempos de falhas e a distribuição Weibull para os tempos de censuras:  $(n = 20; exp(1/500); Weibull(0.9; 500); r = 0.10)$ ,  $(n = 50; exp(1/500); Weibull(0.9; 20); r = 0.10)$  e  $(n = 20; exp(1/2000); Weibull(0.9; 40); r = 0.10)$ , observamos pelas Tabelas B 10.1 a 12.3, em média, um melhor desempenho, em relação ao vício e erro quadrático médio do estimador  $km$ . Porém, entre os estimadores exponenciais por partes, observamos, em média, melhor desempenho do estimador proposto  $mk$ . Os intervalos de confiança bootstrap, das estimativas de sobrevivência calculados nos decis da distribuição teórica, apresentaram probabilidades de cobertura que confirmaram estes resultados.

Para a configuração onde os tempos de falhas e de censuras foram supostos com distribuição Weibull, com a configuração  $(n = 20; Weibull(0.9; 40); Weibull(0.9; 500); r = 0.30)$ , observamos pela Tabela A13.1 a A13.3, melhor desempenho do estimador  $mk$ . Ainda, os estimadores  $ep$  apresentaram em média resultados das estimativas do vício bem menores em relação aos estimadores  $ea$ ,  $km$  e  $en$ . Nesta configuração, o estimador proposto  $mk$  apresentou melhor desempenho em média, em relação ao vício e erro médio quadrático.

Para as situações onde os dados foram perturbados as simulações mostraram que o estimador  $ek$  foi o que apresentou menor alteração, isto é foi o mais resistente por apresentar, de um modo geral, melhor comportamento, para as configurações e perturbações usadas, ou seja, nas configurações:  $(n = 20; exp(1/500); r = 0.30)$  e  $(n = 20; exp(1/2000); r = 0.30)$ , na Perturbação Tipo I. Neste contexto, o segundo mais resistente foi o  $ea$ , até mais da metade da distribuição (6º decil), depois, o estimador  $km$  apresentou-se melhor. Nesta situação o estimador proposto  $mk$  foi o menos resistente. Para a perturbação Tipo II o estimador  $mk$  apresentou-se melhor depois da metade da distribuição, nas seguintes configurações estudadas:  $(n = 20; exp(1/500); r = 0.30)$ ;  $(n = 20; exp(1/40); r = 0.30)$  e  $(n = 50; exp(1/500); r = 0.10)$ . Sob esta perturbação o estimador  $km$  foi o menos resistente.

Ressaltamos que os resultados indicaram, em ambas as perturbações Tipo I e II, o bom desempenho do estimador proposto, nesta dissertação, modificado de Kitchin ( $mk$ ), o qual apresentou em todas as situações menor vício e erro médio quadrático, em

média.

No próximo Capítulo apresentamos as conclusões e perspectivas futuras.

# Capítulo 7

## Conclusões e Perspectivas Futuras

Nesta dissertação propomos um novo estimador do tipo exponencial por partes de Kitchin, que denominamos modificado de Kitchin ( $mk$ ). Verificamos qual estimador apresenta melhor desempenho dentre os mais utilizados. Também nos propomos estudar a vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier e verificada por nós para os demais estimadores, a pequenas perturbações provocadas no conjunto de dados. Neste contexto verificamos qual estimador foi mais resistente às perturbações Tipo I e II entre os estudados.

De acordo com os resultados do estudo de simulação desenvolvido, para as situações específicas em que os tempos de falhas são supostos exponencialmente distribuídos, na ausência das perturbações Tipo I e II, observamos, em média, um melhor desempenho, em relação ao vício e erro médio quadrático do estimador proposto, modificado de Kitchin ( $mk$ ). Os intervalos de confiança bootstrap, das estimativas de sobrevivência, determinadas para os decís da distribuição teórica, apresentaram probabilidades de cobertura que confirmaram estes resultados. Segundo os métodos descritos no Capítulo 5, com as distribuições  $Fe$  e  $Fc$  pertencentes às famílias exponencial e Weibull, na ausência das perturbações, os resultados também indicaram em média melhor desempenho, do estimador proposto  $mk$  em relação ao vício e erro médio quadrático.

Os resultados para as configurações, envolvendo a distribuição exponencial para os tempos de falhas e a distribuição Weibull para os tempos de censuras, observamos, em média, um melhor desempenho, em relação ao vício e erro quadrático médio do estimador  $km$ . Porém, entre os estimadores exponenciais por partes, observamos, em média, me-

lhor desempenho do estimador proposto  $mk$ . Os intervalos de confiança bootstrap, das estimativas de sobrevivência calculados nos decis da distribuição teórica, apresentaram probabilidades de cobertura que confirmaram estes resultados.

Na configuração onde os tempos de falhas e de censuras foram supostos com distribuição Weibull, observamos um melhor desempenho do estimador proposto  $mk$ . Ainda, os estimadores  $ep$  apresentaram em média resultados das estimativas do vício bem menores em relação aos estimadores  $ea$ ,  $km$  e  $en$ . Nesta configuração, o estimador proposto  $mk$  apresentou melhor desempenho em média, em relação ao vício e erro médio quadrático.

Nas situações onde os dados foram perturbados as simulações mostraram que o estimador  $ek$  foi o que apresentou menor alteração, isto é foi o mais resistente por apresentar, de um modo geral, melhor comportamento, para as configurações e perturbações usadas. Para a perturbação Tipo I, o segundo mais resistente foi o  $ea$ , até mais da metade da distribuição (6º decil), depois, o estimador  $km$  apresentou-se melhor. Nesta situação o estimador proposto  $mk$  foi o menos resistente. Devemos ressaltar que um dos objetivos deste trabalho foi estudar o problema da vulnerabilidade já comprovada por Oakes (1993) do estimador de Kaplan-Meier, e verificada por nós para os demais estimadores estudados, a pequenas perturbações provocadas no conjunto de dados. Neste contexto, observamos que para a perturbação Tipo I, o estimador de Kitchin foi o mais resistente. Desta forma apresentamos uma solução para problema que Oakes (1993) cita em seu artigo.

Para a perturbação Tipo II o estimador proposto  $mk$  apresentou-se melhor depois da metade da distribuição, nas as configurações estudadas. Sob esta perturbação o estimador  $km$  foi o menos resistente.

Concluimos que a introdução das perturbações não alteraram significativamente o bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

O estudo de simulação apresentado nesta dissertação considera uma única perturbação do tipo I ou II, entretanto o aumento na quantidade de tais perturbações é de grande interesse para a área e será objeto de estudo posterior.

# Apêndice A

## Simulação

A.1 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ;  $\exp(1/500)$ ;  $r = 0.30$ ).

Tabela A 1.1 Vício, em média.

$n = 20; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0307	0.0300	0.0337	-0.0509	-0.0489
$d_2$	0.0590	0.0576	0.0653	0.0007	0.0039
$d_3$	0.0844	0.0823	0.0943	0.0177	0.0097
$d_4$	0.1065	0.1036	0.1203	0.0439	0.0168
$d_5$	0.1245	0.1208	0.1428	0.0698	0.0228
$d_6$	0.1375	0.1328	0.1609	0.0919	0.0279
$d_7$	0.1438	0.1380	0.1734	0.1106	0.0343
$d_8$	0.1406	0.1337	0.1782	0.1228	0.0402
$d_9$	0.1222	0.1139	0.1716	0.1273	0.0478

Tabela A 1.2 Erro Quadrático médio, em média.

$n = 20; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0009	0.0009	0.0011	0.0081	0.0078
$d_2$	0.0039	0.0038	0.0047	0.0001	0.0001
$d_3$	0.0086	0.0082	0.0102	0.0014	0.0004
$d_4$	0.0121	0.0115	0.0151	0.0045	0.0008
$d_5$	0.0174	0.0165	0.0221	0.0087	0.0014
$d_6$	0.0218	0.0205	0.0286	0.0125	0.0016
$d_7$	0.0222	0.0205	0.0313	0.0171	0.0023
$d_8$	0.0221	0.0201	0.0338	0.0208	0.0030
$d_9$	0.0165	0.0144	0.0309	0.0221	0.0037

Tabela A 1.3 Probabilidade de cobertura,

obtidas via simulação para

os intervalos de confiança

95% (IC) bootstrap, com

a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	100	100	100
$d_6$	100	100	99.8	100	100
$d_7$	100	100	58.0	100	100
$d_8$	100	100	0.0	100	100
$d_9$	100	100	0.0	100	100
$d_{10}$	100	100	0.0	100	100

A.2 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/500)$ ;  $r = 0.10$ ).

Tabela A 2.1 Vício, em média.

$n = 50; \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0105	0.0101	0.0147	-0.0500	-0.0507
$d_2$	0.0200	0.0192	0.0285	0.0046	0.0019
$d_3$	0.0283	0.0269	0.0412	0.0142	0.0045
$d_4$	0.0352	0.0334	0.0527	0.0241	0.0085
$d_5$	0.0406	0.0383	0.0628	0.0317	0.0113
$d_6$	0.0440	0.0413	0.0713	0.0368	0.0166
$d_7$	0.0450	0.0418	0.0777	0.0415	0.0193
$d_8$	0.0428	0.0390	0.0815	0.0431	0.0221
$d_9$	0.0353	0.0310	0.0814	0.0415	0.0253

Tabela A 2.2 Erro quadrático médio, em média.

$n = 50; \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0002	0.0001	0.0003	0.0079	0.0083
$d_2$	0.0005	0.0005	0.0009	0.0001	0.0000
$d_3$	0.0010	0.0009	0.0019	0.0003	0.0001
$d_4$	0.0015	0.0013	0.0030	0.0008	0.0002
$d_5$	0.0020	0.0018	0.0044	0.0014	0.0003
$d_6$	0.0024	0.0021	0.0057	0.0020	0.0005
$d_7$	0.0023	0.0020	0.0063	0.0025	0.0007
$d_8$	0.0020	0.0017	0.0069	0.0025	0.0007
$d_9$	0.0014	0.0011	0.0069	0.0024	0.0009

Tabela A 2.3 Probabilidade de cobertura,  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com  
 a seguinte configuração.

$n = 50; exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	100	100	100
$d_6$	100	100	100	100	100
$d_7$	100	100	100	100	100
$d_8$	100	100	100	100	100
$d_9$	100	100	100	100	100
$d_{10}$	100	100	8.6	100	100

A.3 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/500)$ ;  $r = 0.30$ ).

Tabela A 3.1 Vício, em média.

$n = 50; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0311	0.0301	0.0330	-0.0470	-0.0514
$d_2$	0.0598	0.0577	0.0639	0.0147	0.0017
$d_3$	0.0857	0.0825	0.0921	0.0424	0.0038
$d_4$	0.1085	0.1040	0.1174	0.0689	0.0073
$d_5$	0.1276	0.1217	0.1393	0.0913	0.0097
$d_6$	0.1417	0.1342	0.1568	0.1090	0.0111
$d_7$	0.1495	0.1402	0.1686	0.1230	0.0140
$d_8$	0.1484	0.1368	0.1727	0.1297	0.0161
$d_9$	0.1324	0.1176	0.1648	0.1269	0.0183

Tabela A 3.2 Erro quadrático médio, em média.

$n = 50; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0010	0.0009	0.0011	0.0034	0.0044
$d_2$	0.0036	0.0033	0.0040	0.0004	0.0000
$d_3$	0.0078	0.0072	0.0089	0.0022	0.0001
$d_4$	0.0126	0.0116	0.0145	0.0057	0.0003
$d_5$	0.0177	0.0162	0.0207	0.0100	0.0004
$d_6$	0.0216	0.0196	0.0261	0.0140	0.0004
$d_7$	0.0238	0.0210	0.0297	0.0177	0.0006
$d_8$	0.0237	0.0202	0.0312	0.0193	0.0006
$d_9$	0.0185	0.0146	0.0279	0.0181	0.0007

Tabela A 3.3 Probabilidade de cobertura

obtidas via simulação para

os intervalos de confiança

95% (IC) bootstrap com

a seguinte configuração.

$n = 50; exp(1/500); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	99.5	100	100
$d_4$	99.4	100	16.5	100	100
$d_5$	49.0	91.7	0.0	100	100
$d_6$	2.7	39.0	0.0	100	100
$d_7$	0.1	9.0	0.0	95.2	100
$d_8$	0.0	8.0	0.0	80.2	100
$d_9$	5.3	55.2	0.0	74.0	100

A.4 Desempenho dos estimadores com dados simulados na configuração ( $n = 100$ ;  $\exp(1/500)$ ;  $r = 0.10$ ).

Tabela A 4.1 Vício, em média.

$n = 100; \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0105	0.0101	0.0144	-0.0490	-0.0500
$d_2$	0.0200	0.0191	0.0278	0.0047	0.0008
$d_3$	0.0283	0.0269	0.0401	0.0142	0.0020
$d_4$	0.0352	0.0334	0.0513	0.0235	0.0044
$d_5$	0.0407	0.0384	0.0612	0.0305	0.0059
$d_6$	0.0442	0.0414	0.0693	0.0350	0.0064
$d_7$	0.0453	0.0412	0.0755	0.0390	0.0083
$d_8$	0.0432	0.0392	0.0791	0.0399	0.0096
$d_9$	0.0360	0.0314	0.0789	0.0375	0.0109

Tabela A 4.2 Erro Quadrático médio, em média.

$n = 100; \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0001	0.0001	0.0002	0.0045	0.0047
$d_2$	0.0004	0.0003	0.0007	0.0001	0.0000
$d_3$	0.0008	0.0007	0.0016	0.0003	0.0001
$d_4$	0.0013	0.0011	0.0026	0.0008	0.0001
$d_5$	0.0017	0.0015	0.0038	0.0012	0.0001
$d_6$	0.0020	0.0017	0.0048	0.0016	0.0001
$d_7$	0.0021	0.0018	0.0057	0.0020	0.0002
$d_8$	0.0018	0.0015	0.0061	0.0020	0.0002
$d_9$	0.0013	0.0010	0.0062	0.0018	0.0002

Tabela A 4.3 Probabilidade cobertura obtidas via simulação para os intervalos de confiança 95% (IC) bootstrap com a seguinte configuração.

$n = 100; exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	100	100	100
$d_6$	100	100	100	100	100
$d_7$	100	100	96.0	100	100
$d_8$	100	100	1.9	100	100
$d_9$	100	100	0.0	100	100

A.5 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/2000)$ ;  $r = 0.30$ ).

Tabela A 5.1 Vício, em média.

$n = 50; \exp(1/2000); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0311	0.0301	0.0331	-0.0472	-0.0518
$d_2$	0.0598	0.0577	0.0639	0.0149	0.0016
$d_3$	0.0858	0.0826	0.0922	0.0429	0.0038
$d_4$	0.1086	0.1041	0.1175	0.0686	0.0073
$d_5$	0.1276	0.1217	0.1393	0.0910	0.0098
$d_6$	0.1417	0.1343	0.1568	0.1096	0.0112
$d_7$	0.1496	0.1403	0.1687	0.1229	0.0141
$d_8$	0.1484	0.1367	0.1727	0.1298	0.0163
$d_9$	0.1325	0.1177	0.1648	0.1270	0.0185

Tabela A 5.2 Erro Quadrático médio, em média.

$n = 50; \exp(1/2000); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0011	0.0010	0.0012	0.0063	0.0083
$d_2$	0.0037	0.0035	0.0042	0.0003	0.0000
$d_3$	0.0077	0.0072	0.0088	0.0023	0.0001
$d_4$	0.0123	0.0113	0.0142	0.0056	0.0002
$d_5$	0.0172	0.0157	0.0202	0.0099	0.0003
$d_6$	0.0213	0.0192	0.0257	0.0141	0.0004
$d_7$	0.0229	0.0202	0.0288	0.0175	0.0005
$d_8$	0.0226	0.0193	0.0302	0.0189	0.0005
$d_9$	0.0184	0.0146	0.0277	0.0182	0.0006

Tabela A 5.3 Probabilidade cobertura obtidas via simulação para os intervalos de confiança 95% (IC) bootstrap com a seguinte configuração.

$n = 50; exp(1/2000); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	99.2	100	100
$d_4$	99.5	100	15.7	100	100
$d_5$	48.8	93.0	0.0	100	100
$d_6$	2.0	36.8	0.0	100	100
$d_7$	0.2	7.7	0.0	99.2	100
$d_8$	0.1	7.4	0.0	89.3	100
$d_9$	4.1	53.8	0.0	83.2	100

A.6 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ;  $\exp(1/40)$ ;  $r = 0.30$ ).

Tabela A 6.1 Vício, em média.

$n = 20$ ; $\exp(1/40)$ ; $r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0306	0.0300	0.0337	-0.0559	-0.0535
$d_2$	0.0590	0.0577	0.0653	0.0057	0.0090
$d_3$	0.0845	0.0824	0.0943	0.0168	0.0089
$d_4$	0.1063	0.1034	0.1201	0.0468	0.0206
$d_5$	0.1245	0.1208	0.1428	0.0742	0.0284
$d_6$	0.1376	0.1329	0.1609	0.0894	0.0250
$d_7$	0.1437	0.1380	0.1733	0.1076	0.0309
$d_8$	0.1407	0.1337	0.1782	0.1239	0.0414
$d_9$	0.1220	0.1138	0.1715	0.1293	0.0496

Tabela A 6.2 Erro Quadrático médio, em média.

$n = 20$ ; ; $\exp(1/40)$ ; $r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0010	0.0010	0.0012	0.0072	0.0075
$d_2$	0.0043	0.0042	0.0051	0.0004	0.0002
$d_3$	0.0077	0.0074	0.0095	0.0013	0.0004
$d_4$	0.0117	0.0111	0.0148	0.0049	0.0012
$d_5$	0.0160	0.0151	0.0207	0.0092	0.0019
$d_6$	0.0194	0.0182	0.0263	0.0122	0.0016
$d_7$	0.0230	0.0213	0.0322	0.0163	0.0021
$d_8$	0.0233	0.0213	0.0354	0.0213	0.0034
$d_9$	0.0187	0.0165	0.0337	0.0226	0.0040

Tabela A 6.3 Probabilidade de cobertura  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com a  
 seguinte configuração.

$n = 20; exp(1/40); r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	99.4	100	100
$d_6$	100	100	50.6	100	100
$d_7$	100	100	0.6	100	100
$d_8$	99.8	99.8	0.0	100	100
$d_9$	100	100	0.0	100	100

A.7 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/40)$ ;  $r = 0.10$ ).

Tabela A 7.1 Vício, em média.

$n = 50; \exp(1/40); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0106	0.0101	0.0147	-0.0545	-0.0553
$d_2$	0.0202	0.0193	0.0286	0.0097	0.0073
$d_3$	0.0283	0.0270	0.0412	0.0135	0.0037
$d_4$	0.0352	0.0334	0.0527	0.0280	0.0125
$d_5$	0.0406	0.0383	0.0628	0.0374	0.0172
$d_6$	0.0439	0.0412	0.0712	0.0340	0.0103
$d_7$	0.0449	0.0417	0.0776	0.0380	0.0131
$d_8$	0.0428	0.0390	0.0815	0.0441	0.0203
$d_9$	0.0354	0.0311	0.0815	0.0434	0.0239

Tabela A 7.2 Erro Quadrático médio, em média.

$n = 50; \exp(1/40); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0001	0.0001	0.0002	0.0045	0.0046
$d_2$	0.0005	0.0005	0.0010	0.0002	0.0001
$d_3$	0.0010	0.0010	0.0020	0.0004	0.0001
$d_4$	0.0013	0.0011	0.0028	0.0013	0.0004
$d_5$	0.0016	0.0015	0.0039	0.0022	0.0007
$d_6$	0.0020	0.0018	0.0051	0.0019	0.0004
$d_7$	0.0018	0.0016	0.0057	0.0022	0.0005
$d_8$	0.0017	0.0014	0.0064	0.0029	0.0008
$d_9$	0.0014	0.0011	0.0067	0.0026	0.0010

Tabela A 7.3 Probabilidade de cobertura  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com a  
 seguinte configuração.

$n = 50; \exp(1/40); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	100	100	100
$d_6$	100	100	100	100	100
$d_7$	100	100	100	100	100
$d_8$	100	100	100	100	100
$d_9$	100	100	11.8	100	100

A.8 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ;  $\exp(1/40)$ ;  $\exp(1/500)$ ;  $r = 0.10$ ).

Tabela A 8.1 Vício, em média.

$n = 20; \exp(1/40); \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0429	0.0414	0.0446	-0.0887	-0.0864
$d_2$	0.0771	0.0741	0.0813	-0.0770	-0.0712
$d_3$	0.0673	0.0647	0.0789	-0.0651	-0.0559
$d_4$	0.0576	0.0553	0.0766	-0.0532	-0.0406
$d_5$	0.0480	0.0462	0.0744	-0.0415	-0.0254
$d_6$	0.0384	0.0369	0.0721	-0.0299	-0.0105
$d_7$	0.0287	0.0275	0.0696	-0.0188	0.0039
$d_8$	0.0191	0.0184	0.0671	0.0082	0.0176
$d_9$	0.0095	0.0091	0.0642	0.0014	0.0300

Tabela A 8.2 Erro Quadrático médio, em média.

$n = 20; \exp(1/40); \exp(1/500); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0024	0.0023	0.0026	0.0073	0.0071
$d_2$	0.0075	0.0071	0.0081	0.0047	0.0042
$d_3$	0.0059	0.0056	0.0077	0.0029	0.0021
$d_4$	0.0046	0.0044	0.0074	0.0016	0.0008
$d_5$	0.0036	0.0034	0.0072	0.0008	0.0002
$d_6$	0.0026	0.0025	0.0069	0.0003	0.0000
$d_7$	0.0016	0.0015	0.0063	0.0001	0.0002
$d_8$	0.0006	0.0006	0.0052	0.0000	0.0007
$d_9$	0.0001	0.0001	0.0044	0.0001	0.0015

Tabela A 8.3 Probabilidade de cobertura  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com a  
 seguinte configuração.

$n = 20; \exp(1/40); \exp(1/500);$					
$r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	100	100	100
$d_4$	100	100	100	100	100
$d_5$	100	100	100	100	100
$d_6$	100	100	100	100	100
$d_7$	100	100	100	100	100
$d_8$	100	100	100	100	100
$d_9$	100	100	79	100	100

A.9 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/40)$ ;  $\exp(1/500)$ ;  $r = 0.30$ ).

Tabela A 9.1 Vício, em média.

$n = 50$ ; $\exp(1/40)$ ; $\exp(1/500)$ ; $r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0319	0.0306	0.0336	-0.0718	-0.0709
$d_2$	0.0609	0.0582	0.0644	-0.0415	0.0416
$d_3$	0.0863	0.0821	0.0918	-0.0113	0.0127
$d_4$	0.1076	0.1017	0.1154	0.0188	0.0159
$d_5$	0.1238	0.1158	0.1342	0.0485	0.0439
$d_6$	0.1328	0.1224	0.1465	0.0777	0.0711
$d_7$	0.1316	0.1186	0.1501	0.1060	0.0968
$d_8$	0.1131	0.0992	0.1412	0.1332	0.1203
$d_9$	0.0704	0.0599	0.1175	0.1583	0.1371

Tabela A 9.2 Erro Quadrático médio, em média.

$n = 50$ ; $\exp(1/40)$ ; $\exp(1/500)$ ; $r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0011	0.0010	0.0012	0.0050	0.0048
$d_2$	0.0039	0.0036	0.0043	0.0016	0.0016
$d_3$	0.0081	0.0074	0.0090	0.0001	0.0001
$d_4$	0.0135	0.0122	0.0152	0.0006	0.0005
$d_5$	0.0176	0.0156	0.0202	0.0028	0.0023
$d_6$	0.0199	0.0172	0.0238	0.0071	0.0061
$d_7$	0.0201	0.0168	0.0254	0.0128	0.0110
$d_8$	0.0153	0.0120	0.0224	0.0194	0.0160
$d_9$	0.0062	0.0046	0.0153	0.0265	0.0204

Tabela A 9.3 Probabilidade de cobertura

obtidas via simulação para  
os intervalos de confiança  
95% (IC) bootstrap com a  
seguinte configuração.

$n = 50; exp(1/40); exp(1/500);$					
$r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	96.8	100	100
$d_4$	98.2	100	37.6	100	100
$d_5$	75.4	98.4	1.8	100	100
$d_6$	53.4	97.4	0.0	100	100
$d_7$	78.0	100	0.0	44.8	99.8
$d_8$	100	100	0.2	0.0	30.4
$d_9$	100	100	0.0	0.0	20.3

A.10 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ;  $\exp(1/500)$ ; Weibull (0.9; 500);  $r = 0.10$ ).

Tabela A 10.1 Vício, em média.

$n = 20$ ; $\exp(1/500)$ ; Weibull (0.9; 500);					
$r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0116	0.0113	0.0163	-0.1031	-0.0974
$d_2$	0.0209	0.0203	0.0306	-0.0962	-0.0911
$d_3$	0.0285	0.0276	0.0434	-0.0856	-0.0841
$d_4$	0.0347	0.0335	0.0548	-0.0752	-0.0766
$d_5$	0.0392	0.0378	0.0649	-0.0661	-0.0658
$d_6$	0.0415	0.0398	0.0731	-0.0587	-0.0510
$d_7$	0.0418	0.0399	0.0795	-0.0534	-0.0531
$d_8$	0.0389	0.0368	0.0834	-0.0505	-0.0451
$d_9$	0.0315	0.0292	0.0837	-0.0509	-0.0368

Tabela A 10.2 Erro Quadrático médio, em média.

---



---

$n = 20; \exp(1/500); Weibull(0.9; 500);$					
$r = 0.10$					
$decis$	$ea$	$km$	$en$	$ek$	$mk$
$d_1$	0.0002	0.0002	0.0003	0.0084	0.0086
$d_2$	0.0008	0.0007	0.0013	0.0061	0.0065
$d_3$	0.0013	0.0012	0.0024	0.0036	0.0044
$d_4$	0.0017	0.0017	0.0037	0.0024	0.0033
$d_5$	0.0023	0.0022	0.0052	0.0017	0.0024
$d_6$	0.0022	0.0021	0.0060	0.0014	0.0019
$d_7$	0.0023	0.0021	0.0071	0.0010	0.0013
$d_8$	0.0021	0.0019	0.0079	0.0012	0.0012
$d_9$	0.0013	0.0011	0.0076	0.0019	0.0010

---



---

Tabela A 10.3 Probabilidade de cobertura  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com  
 a seguinte configuração.

---



---

<i>n</i> = 20; <i>exp</i> (1/500); <i>Weibull</i>					
(0.9; 500); <i>r</i> = 0.10					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
<i>d</i> <sub>1</sub>	100	100	100	100	100
<i>d</i> <sub>2</sub>	100	100	100	100	100
<i>d</i> <sub>3</sub>	100	100	100	100	100
<i>d</i> <sub>4</sub>	100	100	100	100	100
<i>d</i> <sub>5</sub>	100	100	100	100	100
<i>d</i> <sub>6</sub>	100	100	100	100	100
<i>d</i> <sub>7</sub>	100	100	100	100	100
<i>d</i> <sub>8</sub>	100	100	100	100	100
<i>d</i> <sub>9</sub>	100	100	100	100	100

---



---

A.11 Desempenho dos estimadores com dados simulados na configuração ( $n = 50$ ;  $\exp(1/500)$ ; Weibull (0.9; 20);  $r = 0.10$ ).

Tabela A 11.1 Vício, em média.

$n = 50$ ; $\exp(1/500)$ ; Weibull (0.9; 20); $r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0120	0.0110	0.0155	-0.2400	-0.1761
$d_2$	0.0216	0.0195	0.0289	-0.2231	-0.1717
$d_3$	0.0283	0.0252	0.0397	-0.2064	-0.1599
$d_4$	0.0314	0.0275	0.0476	-0.1899	-0.1472
$d_5$	0.0302	0.0260	0.0522	-0.1733	-0.1344
$d_6$	0.0256	0.0220	0.0545	-0.1568	-0.1217
$d_7$	0.0195	0.0167	0.0558	-0.1403	-0.1091
$d_8$	0.0128	0.0109	0.0567	-0.1238	-0.0963
$d_9$	0.0063	0.0054	0.0576	-0.1071	-0.0836

Tabela A 11.2 Erro Quadrático médio, em média.

$n = 50$ ; $\exp(1/500)$ ; Weibull (0.9; 20); $r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0002	0.0002	0.0003	0.0397	0.0309
$d_2$	0.0007	0.0006	0.0011	0.0361	0.0291
$d_3$	0.0011	0.0009	0.0020	0.0321	0.0243
$d_4$	0.0015	0.0012	0.0029	0.0287	0.0203
$d_5$	0.0014	0.0011	0.0034	0.0247	0.0166
$d_6$	0.0009	0.0007	0.0034	0.0209	0.0155
$d_7$	0.0006	0.0005	0.0036	0.0174	0.0108
$d_8$	0.0003	0.0002	0.0036	0.0141	0.0084
$d_9$	0.0001	0.0001	0.0036	0.0107	0.0063

Tabela A 11.3 Probabilidade de cobertura  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com a  
 seguinte configuração.

$n = 50; exp(1/500); Weibull$					
$(0.9; 20); r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	10.0	20.1
$d_3$	100	100	100	57.8	31.3
$d_4$	100	100	100	85.0	56.4
$d_5$	100	100	100	99.0	11.2
$d_6$	100	100	100	98.8	46.8
$d_7$	100	100	100	99.6	66.4
$d_8$	100	100	100	82.0	99.2
$d_9$	100	100	100	6.4	71.8

A 12 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ;  $\exp(1/2000)$ ; Weibull  $(0.9, 40)$ ;  $r = 0.10$ ).

Tabela A 12.1 Vício, em média.

$n = 20$ ; $\exp(1/2000)$ ; Weibull $(0.9, 40)$ ; $r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0396	0.0382	0.0416	-0.3669	-0.1774
$d_2$	0.0689	0.0662	0.0739	-0.3796	-0.1620
$d_3$	0.0677	0.0650	0.0791	-0.3496	-0.1456
$d_4$	0.0591	0.0567	0.0778	-0.3121	-0.1291
$d_5$	0.0493	0.0473	0.0755	-0.2740	-0.1127
$d_6$	0.0396	0.0380	0.0731	-0.2356	-0.0961
$d_7$	0.0297	0.0285	0.0705	-0.1974	-0.0799
$d_8$	0.0198	0.0190	0.0678	-0.1590	-0.0632
$d_9$	0.0098	0.0094	0.0646	-0.1202	-0.0464

Tabela A 12.2 Erro Quadrático médio, em média.

$n = 20$ ; $\exp(1/2000)$ ; Weibull $(0.9, 40)$ ; $r = 0.10$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0020	0.0019	0.0021	0.0909	0.0299
$d_2$	0.0057	0.0054	0.0064	0.1008	0.0230
$d_3$	0.0060	0.0056	0.0077	0.0878	0.0163
$d_4$	0.0047	0.0044	0.0073	0.0730	0.0125
$d_5$	0.0032	0.0030	0.0066	0.0605	0.0094
$d_6$	0.0023	0.0022	0.0065	0.0464	0.0066
$d_7$	0.0012	0.0011	0.0056	0.0340	0.0045
$d_8$	0.0007	0.0007	0.0054	0.0230	0.0028
$d_9$	0.0002	0.0002	0.0045	0.0131	0.0014

Tabela A 12.3 Probabilidade de cobertura,  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com a  
 seguinte onfiguração.

---



---

$n = 20$ ;  $exp(1/2000)$ ; *Weibull*

---

$(0.9, 40)$ ;  $r = 0.10$ .

---

<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	2.4	8.4
$d_3$	100	100	100	60.8	100
$d_4$	100	100	100	97.6	100
$d_5$	100	100	100	99.8	100
$d_6$	100	100	100	100	100
$d_7$	100	100	100	100	100
$d_8$	100	100	100	100	100
$d_9$	100	100	65.2	98.0	100

---



---

A.13 Desempenho dos estimadores com dados simulados na configuração ( $n = 20$ ; Weibull (0.9, 40); Weibull (0.9, 500);  $r = 0.30$ ).

Tabela A 13.1 Vício, em média.

$n = 20$ ; Weibull (0.9; 40); Weibull (0.9, 500);					
$r = 0.30$					
<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0490	0.0478	0.0505	-0.0722	-0.0699
$d_2$	0.0937	0.0911	0.0969	-0.0428	-0.0388
$d_3$	0.1324	0.1284	0.1379	-0.0122	-0.0077
$d_4$	0.1628	0.1572	0.1714	0.0188	0.0130
$d_5$	0.1790	0.1720	0.1924	0.0495	0.0430
$d_6$	0.1736	0.1661	0.1950	0.0794	0.0719
$d_7$	0.1461	0.1393	0.1792	0.1083	0.1061
$d_8$	0.1033	0.0982	0.1511	0.1360	0.1339
$d_9$	0.0532	0.0505	0.1171	0.1613	0.1540

Tabela A 13.2 Erro Quadrático médio, em média.

---



---

$n = 20$ ; *Weibull* (0.9, 40); *Weibull* (0.9, 500);

---

$r = 0.30$

---

<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	0.0026	0.0025	0.0028	0.0042	0.0042
$d_2$	0.0096	0.0091	0.0102	0.0011	0.0009
$d_3$	0.0208	0.0198	0.0222	0.0000	0.0000
$d_4$	0.0308	0.0290	0.0334	0.0010	0.0003
$d_5$	0.0369	0.0345	0.0418	0.0036	0.0030
$d_6$	0.0351	0.0325	0.0427	0.0072	0.0067
$d_7$	0.0259	0.0238	0.0366	0.0122	0.0026
$d_8$	0.0132	0.0121	0.0258	0.0182	0.0178
$d_9$	0.0038	0.0035	0.0155	0.0254	0.0227

---



---

Tabela A 13.3 Probabilidade de cobertura,  
 obtidas via simulação para  
 os intervalos de confiança  
 95% (IC) bootstrap com  
 a seguinte configuração.

---



---

$n = 20$ ; *Weibull* (0.9, 40);

---

*Weibull* (0.9, 500);  $r = 0.30$

---

<i>decis</i>	<i>ea</i>	<i>km</i>	<i>en</i>	<i>ek</i>	<i>mk</i>
$d_1$	100	100	100	100	100
$d_2$	100	100	100	100	100
$d_3$	100	100	99.8	100	100
$d_4$	97.2	99.6	58.4	100	100
$d_5$	87.4	98.2	2.8	100	100
$d_6$	94.6	99.6	0.6	100	100
$d_7$	100	100	0.0	100	100
$d_8$	100	100	0.0	75	99.6
$d_9$	100	100	0.0	0.0	100

---



---

# Apêndice B

## Simulação

B.1 Desempenho dos estimadores, segundo os dados simulados, na situação normal, comparados com os dados sujeitos à perturbação do Tipo I (indicada pelo símbolo  $'$ ), na configuração ( $n = 20$ ;  $\exp(1/500)$ ;  $r = 0.30$ ).

Tabela B 1.1 Estimativas, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.9307	0.9405	0.9300	0.9400	0.9337	0.9433	0.9491	0.9463	0.9511	0.9135
$d_2$	0.8590	0.8729	0.8576	0.8729	0.8653	0.8792	0.9007	0.9008	0.9039	0.8276
$d_3$	0.7844	0.7991	0.7823	0.8001	0.7943	0.8095	0.8177	0.8177	0.8097	0.7422
$d_4$	0.7065	0.7200	0.7036	0.7225	0.7203	0.7352	0.7439	0.7440	0.7168	0.6569
$d_5$	0.6245	0.6358	0.6208	0.6402	0.6428	0.6564	0.6698	0.6698	0.6228	0.5983
$d_6$	0.5375	0.5340	0.5328	0.5403	0.5609	0.5612	0.5919	0.5919	0.5279	0.4995
$d_7$	0.4438	0.4337	0.4380	0.4395	0.4734	0.4675	0.5106	0.5106	0.4343	0.4282
$d_8$	0.3406	0.3293	0.3337	0.3328	0.3782	0.3705	0.4228	0.4228	0.3402	0.3392
$d_9$	0.2222	0.2128	0.2139	0.2124	0.2716	0.2646	0.3273	0.3272	0.2478	0.2398

Tabela B 1.2 Vício, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0307	0.0405	0.0300	0.0400	0.0337	0.0433	-0.0509	-0.0537	-0.0489	-0.0865
$d_2$	0.0590	0.0729	0.0576	0.0729	0.0653	0.0792	0.0007	0.0008	0.0039	-0.0724
$d_3$	0.0844	0.0991	0.0823	0.1001	0.0943	0.1095	0.0177	0.0177	0.0097	-0.0578
$d_4$	0.1065	0.1200	0.1036	0.1225	0.1203	0.1352	0.0439	0.0440	0.0168	0.0431
$d_5$	0.1245	0.1358	0.1208	0.1402	0.1428	0.1564	0.0698	0.0698	0.0228	0.0187
$d_6$	0.1375	0.1340	0.1328	0.1403	0.1609	0.1612	0.0919	0.0919	0.0279	0.0040
$d_7$	0.1438	0.1337	0.1380	0.1395	0.1734	0.1675	0.1106	0.1106	0.0343	0.0012
$d_8$	0.1406	0.1293	0.1337	0.1328	0.1782	0.1705	0.1228	0.1228	0.0402	0.0072
$d_9$	0.1222	0.1128	0.1139	0.1124	0.1716	0.1646	0.1273	0.1272	0.0478	0.0262

Tabela B 1.3 Erro Quadrático médio, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0009	0.0018	0.0009	0.0018	0.0011	0.0021	0.0081	0.0077	0.0078	0.0062
$d_2$	0.0039	0.0063	0.0038	0.0063	0.0047	0.0072	0.0001	0.0002	0.0001	0.0033
$d_3$	0.0086	0.0108	0.0082	0.0110	0.0102	0.0129	0.0014	0.0014	0.0004	0.0013
$d_4$	0.0121	0.0150	0.0115	0.0156	0.0151	0.0188	0.0045	0.0043	0.0008	0.0004
$d_5$	0.0174	0.0212	0.0165	0.0224	0.0221	0.0270	0.0087	0.0082	0.0014	0.0000
$d_6$	0.0278	0.0211	0.0205	0.0229	0.0286	0.0290	0.0125	0.0224	0.0016	0.0004
$d_7$	0.0222	0.0203	0.0205	0.0218	0.0313	0.0302	0.0171	0.0176	0.0023	0.0012
$d_8$	0.0221	0.0183	0.0201	0.0192	0.0338	0.0304	0.0208	0.0220	0.0030	0.0022
$d_9$	0.0165	0.0139	0.0144	0.0138	0.0309	0.0282	0.0221	0.0223	0.0037	0.0036

Tabela B 1.4 Probabilidade de cobertura, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, obtidas na simulação para os intervalos de confiança 95% (IC) bootstrap, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	100	100	100	100	100	100	100	100	100	100
$d_2$	100	100	100	100	100	100	100	100	100	100
$d_3$	100	100	100	100	100	100	100	100	100	100
$d_4$	100	100	100	100	100	100	100	100	100	100
$d_5$	100	100	100	100	99.8	99.4	100	100	100	100
$d_6$	100	100	100	100	58.0	76.1	100	100	100	100
$d_7$	100	100	100	100	0.0	2.1	100	100	100	100
$d_8$	100	100	100	100	0.0	0.0	100	100	100	100
$d_9$	100	100	100	100	0.0	0.0	100	100	100	100

Diferenças relativas das estimativas, para a configuração ( $n = 20$ ;  $\exp(1/500)$ ;  $r = 0.30$ ), com a perturbação Tipo I, em relação à situação normal, referentes a Tabela B 1.1.

Tabela B 1.5 Diferenças relativas das estimativas, com a perturbação Tipo I, em relação à situação normal.

$n = 20; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>DFR ea</i>	<i>DFR km</i>	<i>DFR en</i>	<i>DFR ek</i>	<i>DFR mk</i>
$d_1$	-0.0105	-0.0107	-0.0102	0.0029	0.0395
$d_2$	-0.0161	-0.0117	-0.0160	-0.0001	0.0844
$d_3$	-0.0187	-0.0227	-0.0191	0.0000	0.0833
$d_4$	-0.0191	-0.0268	-0.0206	-0.0001	0.0835
$d_5$	-0.0180	-0.0312	-0.0211	0.0000	0.0393
$d_6$	0.0065	-0.0140	-0.0005	0.0000	0.0537
$d_7$	0.0227	-0.0034	0.0124	0.0000	0.0140
$d_8$	0.0331	0.0026	0.0203	0.0000	0.0029
$d_9$	0.0423	0.0070	0.0257	0.0003	0.0322

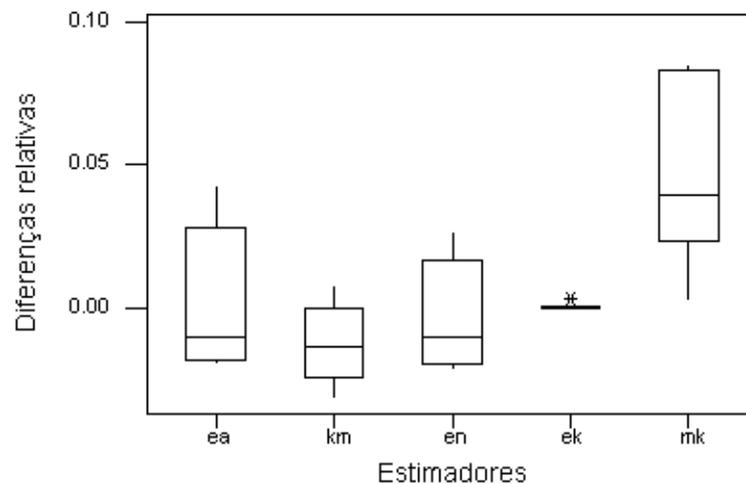


Figura B 1 Desenhos esquemáticos das distribuições das diferenças relativas das estimativas, na configuração considerada com a perturbação Tipo I, em relação à situação normal.

Na Tabela B.1.5 os resultados indicam que o estimador  $ek$  foi o que apresentou menor alteração, isto é foi o mais resistente, pois através do desenho esquemático, apresentado na Figura B 1, o  $ek$  está quase todo concentrado em torno de zero, só existe um ponto atípico maior que zero, para a configuração e perturbação usadas. O segundo melhor foi o  $ea$  até mais da metade da distribuição (6º decil). Depois do 6º decil o segundo melhor foi o  $km$ . O estimador  $mk$  foi o menos resistente.

De um modo geral, observamos na Fig.B 1 que medianamente, os estimadores  $ea$  e  $en$  parecem indicar os mesmos efeitos. Enquanto para os estimadores  $ea$ ,  $en$  e  $mk$  as distribuições foram assimétricas, para o estimador  $km$  as diferenças relativas apresentaram-se simetricamente distribuídas. O estimador  $mk$  foi o que apresentou as maiores alterações.

A introdução da perturbação não alterou o bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

B.2 Desempenho dos estimadores, segundo os dados simulados, na situação normal, comparados com os dados sujeitos à perturbação do Tipo I (indicada pelo símbolo  $'$ ), na configuração ( $n = 20$ ;  $\exp(1/2000)$ ;  $r = 0.30$ ).

Tabela B 2.1 Estimativas, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/2000); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.9308	0.9407	0.9301	0.9401	0.9338	0.9434	0.9475	0.9486	0.9496	0.9135
$d_2$	0.8590	0.8731	0.8577	0.8730	0.8653	0.8793	0.9011	0.9008	0.9043	0.8276
$d_3$	0.7844	0.7990	0.7823	0.8001	0.7943	0.8094	0.8176	0.8178	0.8102	0.7419
$d_4$	0.7066	0.7198	0.7037	0.7223	0.7204	0.7350	0.7431	0.7433	0.7162	0.6866
$d_5$	0.6247	0.6357	0.6209	0.6401	0.6429	0.6563	0.6688	0.6691	0.6220	0.5980
$d_6$	0.5376	0.5338	0.5329	0.5401	0.5610	0.5611	0.5919	0.5920	0.5282	0.4998
$d_7$	0.4439	0.4335	0.4382	0.4393	0.4735	0.4673	0.5101	0.5102	0.4340	0.4280
$d_8$	0.3409	0.3287	0.3339	0.3322	0.3784	0.3701	0.4223	0.4224	0.3400	0.3380
$d_9$	0.2224	0.2125	0.2142	0.2121	0.2719	0.2644	0.3268	0.3270	0.2476	0.2358

Tabela B 2.2 Vício, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/2000); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0308	0.0407	0.0301	0.0401	0.0338	0.0434	-0.0525	-0.0514	-0.0504	-0.0865
$d_2$	0.0590	0.0731	0.0577	0.0730	0.0653	0.0793	0.0011	0.0008	0.0043	-0.0724
$d_3$	0.0844	0.0990	0.0823	0.1001	0.0943	0.1094	0.0176	0.0178	0.0102	-0.0581
$d_4$	0.1066	0.1198	0.1037	0.1223	0.1204	0.1350	0.0431	0.0433	0.0162	-0.0434
$d_5$	0.1247	0.1357	0.1209	0.1401	0.1429	0.1563	0.0688	0.0691	0.0220	-0.0290
$d_6$	0.1376	0.1338	0.1329	0.1401	0.1610	0.1611	0.0919	0.0920	0.0282	-0.0142
$d_7$	0.1439	0.1335	0.1382	0.1393	0.1735	0.1673	0.1101	0.1102	0.0340	-0.0008
$d_8$	0.1409	0.1287	0.1339	0.1322	0.1784	0.1701	0.1293	0.1224	0.0400	0.0170
$d_9$	0.1224	0.1125	0.1142	0.1121	0.1719	0.1644	0.1268	0.1270	0.0476	0.0358

Tabela B 2.3 Erro Quadrático médio, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, com a seguinte configuração.

$n = 20; \exp(1/2000); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0010	0.0020	0.0010	0.0020	0.0012	0.0022	0.0046	0.0036	0.0043	0.0049
$d_2$	0.0037	0.0069	0.0036	0.0069	0.0045	0.0078	0.0002	0.0003	0.0001	0.0017
$d_3$	0.0076	0.0113	0.0073	0.0116	0.0093	0.0134	0.0016	0.0016	0.0005	0.0005
$d_4$	0.0117	0.0158	0.0111	0.0166	0.0147	0.0197	0.0046	0.0048	0.0012	0.0001
$d_5$	0.0162	0.0183	0.0154	0.0195	0.0211	0.0241	0.0083	0.0090	0.0017	0.0000
$d_6$	0.0209	0.0192	0.0196	0.0210	0.0278	0.0271	0.0128	0.0133	0.0022	0.0008
$d_7$	0.0231	0.0191	0.0213	0.0210	0.0322	0.0292	0.0169	0.0182	0.0027	0.0014
$d_8$	0.0217	0.0195	0.0198	0.0207	0.0335	0.0318	0.0194	0.0206	0.0029	0.0023
$d_9$	0.0186	0.0162	0.0164	0.0163	0.0329	0.0306	0.0203	0.0214	0.0031	0.0033

Tabela B 2.4 Probabilidade de cobertura, dos dados não perturbados comparados com os dados sob a perturbação do Tipo I, obtidas na simulação para os intervalos de confiança 95% (IC) bootstrap, com a seguinte configuração

$n = 20; exp(1/2000); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	100	100	100	100	100	100	100	100	100	100
$d_2$	100	100	100	100	100	100	100	100	100	100
$d_3$	100	100	100	100	100	100	100	100	100	100
$d_4$	100	100	100	100	100	100	100	100	100	100
$d_5$	100	100	100	100	98.8	94.2	100	100	100	100
$d_6$	100	100	100	100	51.4	65.2	100	100	100	100
$d_7$	100	100	100	100	0.0	8.0	100	100	100	100
$d_8$	100	100	100	100	0.0	0.0	100	100	100	100
$d_9$	100	100	10	100	0.0	0.0	100	100	100	100

Diferenças relativas das estimativas, para a configuração ( $n = 20$ ;  $\exp(1/2000)$ ;  $r = 0.30$ ), com a perturbação Tipo I, em relação à situação normal, referentes a Tabela B 2.1.

Tabela B 2.5 Diferenças relativas das estimativas, com a perturbação Tipo I, em relação à situação normal.

$n = 20; \exp(1/2000); r = 0.30$					
<i>decis</i>	<i>DFR ea</i>	<i>DFR km</i>	<i>DFR en</i>	<i>DFR ek</i>	<i>DFR mk</i>
$d_1$	-0.0106	-0.0107	-0.0102	-0.0011	0.0380
$d_2$	-0.0164	-0.0178	-0.0161	0.0003	0.0848
$d_3$	-0.0186	-0.0222	-0.0190	-0.0002	0.0843
$d_4$	-0.0186	-0.0264	-0.0202	-0.0002	0.0413
$d_5$	-0.0176	-0.0309	-0.0208	-0.0004	0.0385
$d_6$	0.0070	-0.0135	-0.0001	-0.0001	0.0537
$d_7$	0.0234	-0.0025	0.0130	-0.0001	0.0138
$d_8$	0.0357	0.0050	0.0219	-0.0002	0.0058
$d_9$	0.0445	0.0098	0.0275	-0.0006	0.0476

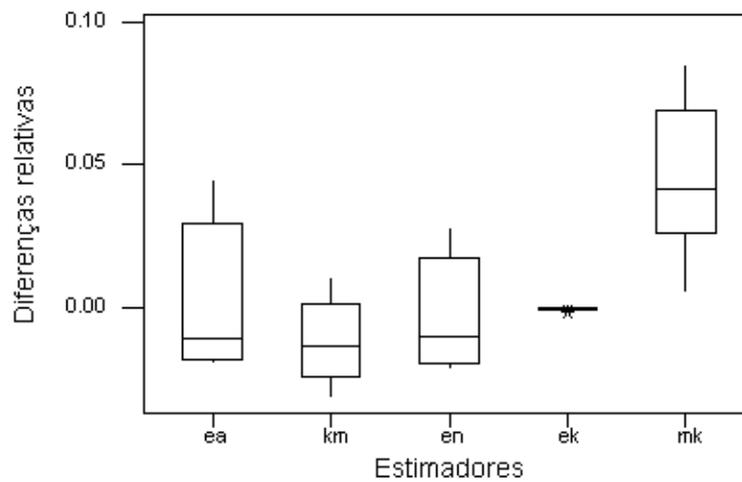


Figura B 2 Desenhos esquemáticos das distribuições das diferenças relativas das estimativas, na configuração considerada com a perturbação Tipo I, em relação à situação normal.

Os resultados da Tabela B.2.5 indicam que, novamente, o estimador *ek* foi o que apresentou menor alteração, isto é foi o mais resistente, para esta outra configuração e mesma perturbação usadas. O segundo melhor foi o *ea* e o *mk* foi o menos resistente. Depois do 6o decil o segundo melhor foi o *km*.

De um modo geral, observamos na Fig.B 2 que medianamente, novamente, os estimadores *ea* e *en* parecem indicar os mesmos efeitos. Enquanto para os estimadores *ea* e *en* as distribuições foram assimétricas, para o estimador *km* as diferenças relativas apresentaram-se simetricamente distribuídas. O estimador *mk* foi o que apresentou as maiores alterações.

A introdução da perturbação não alterou o bom desempenho do estimador proposto *mk*, com relação ao vício e erro médio quadrático, em média.

B.3 Desempenho dos estimadores, segundo os dados simulados, na situação normal, comparados com os dados sujeitos à perturbação do Tipo II (indicada pelo símbolo  $'$ ), na configuração ( $n = 20$ ;  $\exp(1/500)$ ;  $r = 0.30$ ).

Tabela B 3.1 Estimativas, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.9307	0.9405	0.9300	0.9401	0.9337	0.9434	0.9491	0.9493	0.9511	0.9135
$d_2$	0.8590	0.8782	0.8576	0.8784	0.8653	0.8841	0.9007	0.9021	0.9039	0.8278
$d_3$	0.7844	0.8123	0.7823	0.8147	0.7943	0.8218	0.8177	0.8173	0.8097	0.7426
$d_4$	0.7065	0.7425	0.7036	0.7487	0.7203	0.7560	0.7439	0.7400	0.7168	0.6575
$d_5$	0.6245	0.6678	0.6208	0.6802	0.6428	0.6859	0.6698	0.6618	0.6228	0.5984
$d_6$	0.5375	0.5744	0.5328	0.5966	0.5609	0.5985	0.5919	0.5798	0.5279	0.4992
$d_7$	0.4438	0.4740	0.4380	0.5086	0.4734	0.5052	0.5106	0.4950	0.4343	0.4254
$d_8$	0.3406	0.3636	0.3337	0.4151	0.3782	0.4038	0.4228	0.4045	0.3402	0.3398
$d_9$	0.2222	0.2365	0.2139	0.3136	0.2716	0.2901	0.3273	0.3073	0.2478	0.2367

Tabela B 3.2 Vício, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; \exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0307	0.0405	0.0300	0.0401	0.0337	0.0434	-0.0509	-0.0507	-0.0489	-0.0865
$d_2$	0.0590	0.0782	0.0576	0.0784	0.0653	0.0841	0.0007	0.0021	0.0039	-0.0722
$d_3$	0.0844	0.1123	0.0823	0.1147	0.0943	0.1218	0.0177	0.0173	0.0097	-0.0574
$d_4$	0.1065	0.1425	0.1036	0.1487	0.1203	0.1560	0.0439	0.0400	0.0168	-0.0425
$d_5$	0.1245	0.1678	0.1208	0.1802	0.1428	0.1859	0.0698	0.0618	0.0228	-0.0276
$d_6$	0.1375	0.1744	0.1328	0.1966	0.1609	0.1985	0.0919	0.0798	0.0079	-0.0028
$d_7$	0.1438	0.1740	0.1380	0.2086	0.1734	0.2052	0.1106	0.0950	0.0343	0.0024
$d_8$	0.1406	0.1636	0.1337	0.2151	0.1782	0.2038	0.1228	0.1045	0.0402	0.0084
$d_9$	0.1222	0.1365	0.1139	0.2136	0.1716	0.1901	0.1273	0.1073	0.0478	0.0467

Tabela B 3.3 Erro Quadrático médio, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0009	0.0017	0.0009	0.0016	0.0011	0.0019	0.0081	0.0077	0.0078	0.0063
$d_2$	0.0039	0.0072	0.0038	0.0073	0.0047	0.0081	0.0001	0.0001	0.0001	0.0033
$d_3$	0.0086	0.0143	0.0082	0.0149	0.0102	0.0165	0.0014	0.0013	0.0004	0.0013
$d_4$	0.0121	0.0215	0.0115	0.0234	0.0151	0.0254	0.0045	0.0035	0.0008	0.0004
$d_5$	0.0174	0.0305	0.0165	0.0348	0.0221	0.0367	0.0087	0.0065	0.0014	0.0000
$d_6$	0.0218	0.0337	0.0205	0.0419	0.0286	0.0424	0.0125	0.0096	0.0016	0.0003
$d_7$	0.0222	0.0327	0.0205	0.0460	0.0313	0.0142	0.0171	0.0126	0.0023	0.0010
$d_8$	0.0221	0.0294	0.0201	0.0490	0.0338	0.0439	0.0208	0.0153	0.0037	0.0024
$d_9$	0.0165	0.0209	0.0144	0.0480	0.0309	0.0382	0.0221	0.0158	0.0049	0.0036

Tabela B 3.4 Probabilidade de cobertura, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, obtidas na simulação para os intervalos de confiança 95% (IC) bootstrap, com a seguinte configuração

$n = 20; exp(1/500); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	100	100	100	100	100	100	100	100	100	100
$d_2$	100	100	100	100	100	100	100	100	100	100
$d_3$	100	100	100	100	100	100	100	100	100	100
$d_4$	100	100	100	100	100	97.4	100	100	100	100
$d_5$	100	99.5	100	43.4	99.8	3.0	100	100	100	100
$d_6$	100	99.0	100	0.2	58.0	0.0	100	100	100	100
$d_7$	100	97.5	100	0.0	0.0	0.0	100	100	100	100
$d_8$	100	99.5	100	0.0	0.0	0.0	100	100	100	100
$d_9$	100	100	100	0.0	0.0	0.0	100	100	100	100

Diferenças relativas das estimativas, para a configuração ( $n = 20$ ;  $\exp(1/500)$ ;  $r = 0.30$ ), com a perturbação Tipo II, em relação à situação normal, referentes a Tabela B 3.1.

Tabela B 3.5 Diferenças relativas das estimativas, com a perturbação Tipo II, em relação à situação normal.

$n = 20; \exp(1/500); r = 0.30$					
<i>decis</i>	<i>DFR ea</i>	<i>DFR km</i>	<i>DFR en</i>	<i>DFR ek</i>	<i>DFR mk</i>
$d_1$	-0.0105	-0.0108	-0.0103	-0.0002	0.0395
$d_2$	-0.0223	-0.0242	-0.0217	-0.0015	0.0841
$d_3$	-0.0355	-0.0414	-0.0346	0.0004	0.0828
$d_4$	-0.0509	-0.0640	-0.0495	0.0052	0.0827
$d_5$	-0.0693	-0.0956	-0.0670	0.0119	0.0391
$d_6$	-0.0686	-0.0197	-0.0670	0.0204	0.0143
$d_7$	-0.0680	-0.1611	-0.0671	0.0305	0.0204
$d_8$	-0.0675	-0.2439	-0.0676	0.0432	0.0011
$d_9$	-0.0643	-0.4661	-0.0681	0.0611	0.0447

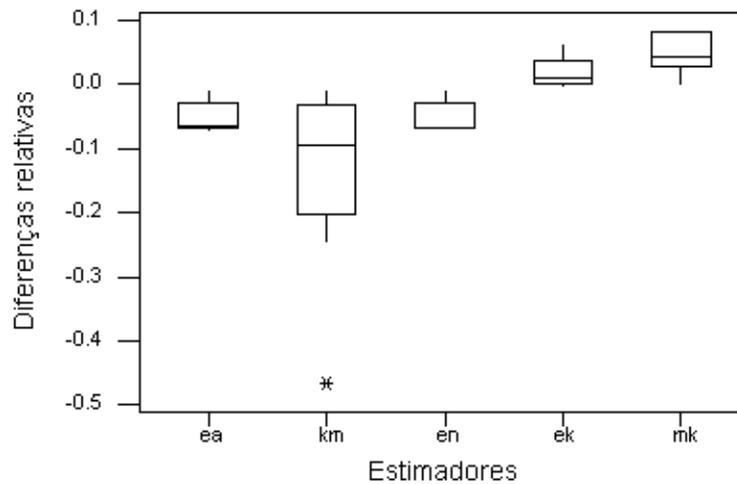


Figura B 3 Desenhos esquemáticos das distribuições das diferenças relativas das estimativas na configuração considerada com a perturbação Tipo II, em relação à situação normal.

Na Tabela B.3.5 os resultados nos indicam que até metade da distribuição (5o decil) o estimador  $ek$  foi o que apresentou menor alteração, isto é foi o mais resistente, para a configuração e perturbação usadas. Depois do 5o decil o melhor foi o estimador proposto  $mk$ . O estimador  $km$  foi o menos resistente.

De um modo geral, observamos na Fig.B 3 que medianamente, os estimadores  $ea$  e  $en$  parecem indicar os mesmos efeitos, nesta configuração e para uma única perturbação Tipo II, como foi observado nas duas situações anteriores, para a uma única perturbação Tipo I. Em todos os estimadores as diferenças relativas apresentaram-se assimetricamente distribuídas. O estimador  $km$  foi o que apresentou as maiores alterações.

Mesmo nessa outra perturbação realizada, a introdução da perturbação não alterou o bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

B.4 Desempenho dos estimadores, segundo os dados simulados, na situação normal, comparados com os dados sujeitos à perturbação do Tipo II (indicada pelo símbolo '), na configuração ( $n = 20$ ;  $\exp(1/40)$ ;  $r = 0.30$ ).

Tabela B 4.1 Estimativas, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; \exp(1/40); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.9306	0.9406	0.9300	0.9402	0.9337	0.9435	0.9438	0.9448	0.9462	0.9114
$d_2$	0.8590	0.8781	0.8577	0.8784	0.8653	0.8841	0.9054	0.9043	0.9089	0.8255
$d_3$	0.7845	0.8123	0.7824	0.8146	0.7943	0.8218	0.8163	0.8168	0.8086	0.7416
$d_4$	0.7063	0.7423	0.7034	0.7485	0.7201	0.7558	0.7463	0.7435	0.7205	0.6767
$d_5$	0.6245	0.6676	0.6208	0.6800	0.6428	0.6857	0.6738	0.6667	0.6283	0.5917
$d_6$	0.5376	0.5743	0.5329	0.5964	0.5609	0.5983	0.5894	0.5773	0.5250	0.4993
$d_7$	0.4437	0.4739	0.4380	0.5085	0.4733	0.5050	0.5075	0.4820	0.4310	0.4225
$d_8$	0.3407	0.3632	0.3337	0.4147	0.3782	0.4035	0.4236	0.4057	0.3411	0.3300
$d_9$	0.2220	0.2363	0.2138	0.3135	0.2715	0.2899	0.3288	0.3002	0.2493	0.2367

Tabela B 4.2 Vício, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; exp(1/40); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0306	0.0406	0.0300	0.0402	0.0337	0.0435	-0.0562	-0.0552	-0.0538	-0.0886
$d_2$	0.0590	0.0781	0.0577	0.0784	0.0653	0.0841	0.0054	0.0043	0.0089	-0.0745
$d_3$	0.0845	0.1123	0.0824	0.1146	0.0943	0.1218	0.0163	0.0168	0.0086	-0.0584
$d_4$	0.1063	0.1423	0.1034	0.1485	0.1201	0.1558	0.0463	0.0435	0.0205	-0.0403
$d_5$	0.1245	0.1676	0.1208	0.1800	0.1428	0.1857	0.0738	0.0667	0.0283	-0.0284
$d_6$	0.1376	0.1743	0.1329	0.1964	0.1609	0.1983	0.0894	0.0773	0.0250	-0.0129
$d_7$	0.1437	0.1739	0.1380	0.2085	0.1733	0.2050	0.1075	0.0920	0.0310	0.0024
$d_8$	0.1407	0.1632	0.1337	0.2147	0.1782	0.2035	0.1236	0.1057	0.0011	0.0085
$d_9$	0.1220	0.1363	0.1138	0.2135	0.1715	0.1899	0.1288	0.1092	0.0493	0.0364

Tabela B 4.3 Erro Quadrático médio, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 20; exp(1/40); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0010	0.0016	0.0010	0.0016	0.0012	0.0018	0.0004	0.0041	0.0045	0.0058
$d_2$	0.0043	0.0070	0.0042	0.0070	0.0051	0.0079	0.0016	0.0004	0.0002	0.0027
$d_3$	0.0077	0.0145	0.0074	0.0150	0.0095	0.0166	0.0052	0.0016	0.0005	0.0006
$d_4$	0.0117	0.0222	0.0111	0.0239	0.0148	0.0260	0.0094	0.0049	0.0015	0.0001
$d_5$	0.0160	0.0313	0.0151	0.0354	0.0207	0.0374	0.0128	0.0090	0.0021	0.0000
$d_6$	0.0194	0.0337	0.0182	0.0415	0.0263	0.0422	0.0168	0.0113	0.0018	0.0004
$d_7$	0.0230	0.0345	0.0213	0.0474	0.0322	0.0462	0.0195	0.0142	0.0022	0.0009
$d_8$	0.0233	0.0298	0.0213	0.0490	0.0354	0.0446	0.0209	0.0169	0.0027	0.0030
$d_9$	0.0187	0.0222	0.0165	0.0491	0.0337	0.0395	0.0181	0.0174	0.0033	0.0049

Tabela B 4.4 Probabilidade de cobertura, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, obtidas na simulação para os intervalos de confiança 95% (IC) bootstrap, com as seguintes configurações.

$n = 20; \exp(1/40); r = 0.30$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	100	100	100	100	100	100	100	100	100	100
$d_2$	100	100	100	100	100	100	100	100	100	100
$d_3$	100	100	100	100	100	100	100	100	100	100
$d_4$	100	100	100	100	100	90.2	100	100	100	100
$d_5$	100	97.4	100	44.8	99.4	6.0	100	100	100	100
$d_6$	100	93.4	100	2.0	50.6	0.0	100	100	100	100
$d_7$	100	87.2	100	0.0	0.6	0.0	100	100	100	100
$d_8$	99.8	96.4	99.8	0.0	0.0	0.0	100	100	100	100
$d_9$	100	100	100	0.0	0.0	0.0	100	100	100	100

Diferenças relativas das estimativas, para a configuração ( $n = 20$ ;  $\exp(1/40)$ ;  $r = 0.30$ ), com a perturbação Tipo II, em relação à situação normal, referentes a Tabela B 4.1.

Tabela B 4.5 Diferenças relativas das estimativas, com a perturbação Tipo II, em relação à situação normal.

$n = 20; \exp(1/40); r = 0.30$					
<i>decis</i>	<i>DFR ea</i>	<i>DFR km</i>	<i>DFR en</i>	<i>DFR ek</i>	<i>DFR mk</i>
$d_1$	-0.0107	-0.0109	-0.0104	-0.0010	0.0367
$d_2$	-0.0222	-0.0241	-0.0217	0.0012	0.0917
$d_3$	-0.0354	-0.0411	-0.0346	-0.0006	0.0828
$d_4$	-0.0509	-0.0641	-0.0495	0.0037	0.0607
$d_5$	-0.0690	-0.0953	-0.0666	0.0105	0.0582
$d_6$	-0.0682	-0.1191	-0.0666	0.0205	0.0289
$d_7$	-0.0680	-0.1609	-0.0669	0.0502	0.0197
$d_8$	-0.0666	-0.2427	-0.0668	0.0422	0.0325
$d_9$	-0.0644	-0.4663	-0.0677	0.0869	0.0505

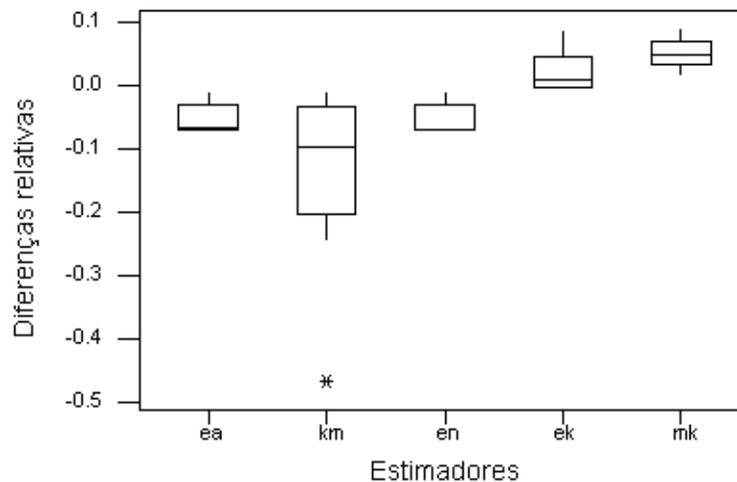


Figura B 4 Desenhos esquemáticos das distribuição das diferenças relativas das estimativas, na configuração considerada com a perturbação Tipo II, em relação à situação normal.

Os resultados da Tabela B.4.1 indicam que até mais da metade da distribuição (6o decil) o estimador  $ek$  foi o mais resistente e o segundo melhor foi o  $ea$ , para a configuração e perturbação usadas. Depois do 6o decil o  $mk$  apresentou-se melhor, em relação aos demais. O menos resistente, novamente, foi o estimador  $km$ .

De um modo geral, observamos na Fig.B 4 que o estimador  $mk$ , nesta configuração e para uma única perturbação Tipo II, foi o que apresentou menor alteração. Medianamente, os estimadores  $ea$ , e  $en$  continuam parecendo apresentar os mesmos resultados. Enquanto para os estimadores  $ea$ ,  $km$ ,  $en$  e  $ek$  as distribuições foram assimétricas, para o estimador  $mk$  as diferenças relativas apresentaram-se simetricamente distribuídas. O estimador  $km$ , novamente, foi o que apresentou a maior alteração.

Os resultados observados, quanto a introdução da perturbação, continuam indicando a não alteração do bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

B.5 Desempenho dos estimadores, segundo os dados simulados na situação normal, comparados com os dados sujeitos à perturbação do Tipo II (indicada pelo símbolo  $'$ ), na configuração ( $n = 50$ ;  $\exp(1/500)$ ;  $r = 0.10$ ).

Tabela B 5.1 Estimativas, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 50; \exp(1/500); r = 0.10$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.9105	0.9144	0.9101	0.9140	0.9147	0.9184	0.9500	0.9503	0.9493	0.9113
$d_2$	0.8200	0.8274	0.8192	0.8274	0.8285	0.8355	0.9046	0.9038	0.9019	0.8230
$d_3$	0.7283	0.7388	0.7269	0.7397	0.7412	0.7512	0.8142	0.8116	0.8045	0.7348
$d_4$	0.6352	0.6484	0.6334	0.6510	0.6527	0.6653	0.7241	0.7199	0.7085	0.6645
$d_5$	0.5406	0.5559	0.5383	0.5613	0.5628	0.5776	0.6317	0.6162	0.6113	0.5881
$d_6$	0.4440	0.4564	0.4413	0.4658	0.4713	0.4836	0.5368	0.5203	0.5132	0.4995
$d_7$	0.3450	0.3544	0.3418	0.3687	0.3777	0.3876	0.4415	0.4241	0.4166	0.3916
$d_8$	0.2428	0.2491	0.2390	0.2692	0.2815	0.2889	0.3431	0.3255	0.3193	0.3031
$d_9$	0.1353	0.1384	0.1310	0.1663	0.1814	0.1862	0.2415	0.2240	0.2221	0.2191

Tabela B 5.2 Vício, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 50; \exp(1/500); r = 0.10$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0105	0.0144	0.0101	0.0140	0.0147	0.0184	-0.0500	-0.0497	-0.0507	-0.0887
$d_2$	0.0200	0.0274	0.0192	0.0273	0.0285	0.0355	0.0046	0.0038	0.0019	-0.0770
$d_3$	0.0283	0.0388	0.0269	0.0397	0.0412	0.0512	0.0142	0.0116	0.0045	-0.0652
$d_4$	0.0352	0.0484	0.0334	0.0510	0.0527	0.0653	0.0241	0.0199	0.0085	-0.0535
$d_5$	0.0406	0.0559	0.0383	0.0613	0.0628	0.0776	0.0371	0.0262	0.0014	-0.0020
$d_6$	0.0440	0.0564	0.0413	0.0658	0.0713	0.0836	0.0368	0.0302	0.0032	-0.0003
$d_7$	0.0450	0.0544	0.0418	0.0687	0.0777	0.0876	0.0415	0.0341	0.0066	-0.0056
$d_8$	0.0428	0.0491	0.0390	0.0692	0.0815	0.0889	0.0431	0.0356	0.0083	-0.0060
$d_9$	0.0353	0.0384	0.0310	0.0663	0.0814	0.0862	0.0415	0.0344	0.0031	0.0021

Tabela B 5.3 Erro Quadrático médio, em média, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, com a seguinte configuração.

$n = 50; \exp(1/500); r = 0.10$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	0.0002	0.0002	0.0001	0.0002	0.0003	0.0003	0.0079	0.0082	0.0083	0.0072
$d_2$	0.0005	0.0009	0.0005	0.0009	0.0009	0.0014	0.0001	0.0000	0.0000	0.0044
$d_3$	0.0010	0.0018	0.0009	0.0019	0.0019	0.0030	0.0003	0.0002	0.0001	0.0024
$d_4$	0.0015	0.0029	0.0013	0.0031	0.0030	0.0048	0.0008	0.0007	0.0002	0.0013
$d_5$	0.0020	0.0039	0.0018	0.0045	0.0044	0.0068	0.0014	0.0011	0.0003	0.0005
$d_6$	0.0024	0.0039	0.0021	0.0051	0.0057	0.0079	0.0020	0.0014	0.0005	0.0000
$d_7$	0.0023	0.0034	0.0020	0.0052	0.0063	0.0082	0.0025	0.0018	0.0007	0.0001
$d_8$	0.0020	0.0028	0.0017	0.0052	0.0069	0.0084	0.0024	0.0018	0.0009	0.0003
$d_9$	0.0014	0.0016	0.0011	0.0046	0.0069	0.0076	0.0015	0.0017	0.0008	0.0005

Tabela B 5.4 Probabilidade de cobertura, dos dados não perturbados comparados com os dados sob a perturbação do Tipo II, obtidas na simulação para os intervalos de confiança 95% (IC) bootstrap, com a seguinte configuração.

$n = 50; \exp(1/500); r = 0.10$										
<i>decis</i>	<i>ea</i>	<i>ea'</i>	<i>km</i>	<i>km'</i>	<i>en</i>	<i>en'</i>	<i>ek</i>	<i>ek'</i>	<i>mk</i>	<i>mk'</i>
$d_1$	100	100	100	100	100	100	100	100	100	100
$d_2$	100	100	100	100	100	100	100	100	100	100
$d_3$	100	100	100	100	100	100	100	100	100	100
$d_4$	100	100	100	100	100	100	100	100	100	100
$d_5$	100	100	100	100	100	100	100	100	100	100
$d_6$	100	100	100	100	100	100	100	100	100	100
$d_7$	100	100	100	100	100	100	100	100	100	100
$d_8$	100	100	100	100	100	100	100	100	100	100
$d_9$	100	100	100	100	8.6	0.6	100	100	100	100

Diferenças relativas das estimativas, para a configuração ( $n = 50$ ;  $\exp(1/500)$ ;  $r = 0.10$ ), com a perturbação Tipo II, em relação à situação normal, referentes a Tabela B 5.1.

Tabela B 5.6 Diferenças relativas das estimativas, com a perturbação Tipo II, em relação à situação normal.

$n = 50; \exp(1/500); r = 0.10$					
<i>decis</i>	<i>DFR ea</i>	<i>DFR km</i>	<i>DFR en</i>	<i>DFR ek</i>	<i>DFR mk</i>
$d_1$	-0.0042	-0.0042	-0.0040	-0.0003	0.0400
$d_2$	0.0090	-0.0100	-0.0084	0.0000	0.0874
$d_3$	-0.0144	-0.0176	-0.0134	0.0031	0.0866
$d_4$	-0.0207	-0.0277	-0.0193	0.0058	0.0621
$d_5$	-0.0283	-0.0427	-0.0262	0.0245	0.0379
$d_6$	-0.0279	-0.0555	-0.0260	0.0307	0.0266
$d_7$	-0.0272	-0.0787	-0.0262	0.0394	0.0200
$d_8$	-0.0259	-0.1263	-0.0262	0.0512	0.0207
$d_9$	-0.0229	-0.2694	-0.0264	0.0724	0.0135

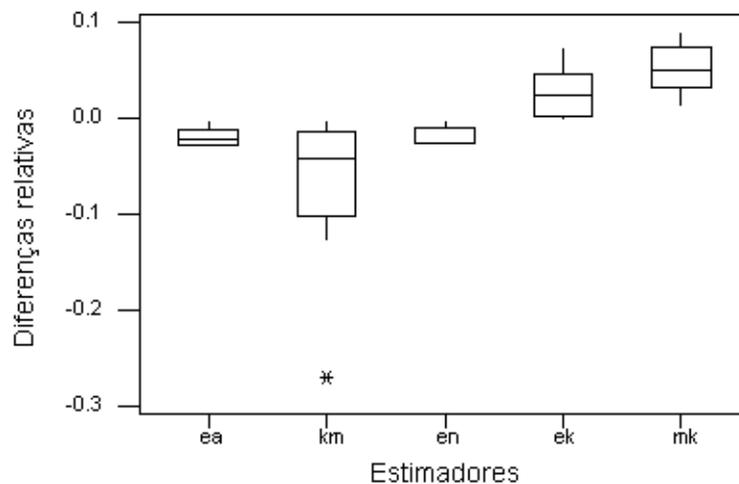


Figura B 5 Desenhos esquemáticos das distribuição das diferenças relativas das estimativas, na configuração considerada com a perturção Tipo II, em relação à situação normal.

Na Tabela B.5.1 os resultados indicam que até mais da metade da distribuição (6o decil) o estimador  $ek$  foi o mais resistente, para a configuração e perturbação usadas. Depois do 6o decil o segundo melhor foi o  $mk$  e depois o  $ea$ . O menos resistente continua sendo o estimador  $km$ .

De um modo geral, observamos na Fig.B 5 que medianamente, os estimadores  $ek$  e  $mk$  parecem indicar os mesmos efeitos. Os estimadores  $km$  e  $en$  apresentaram distribuições assimétricas, para os estimadores  $ea$ ,  $ek$  e  $mk$  as diferenças relativas apresentaram-se simetricamente distribuídas. O estimador  $km$  manteve-se apresentando as maiores alterações.

Mesmo nessa outra configuração e perturbação realizada, a introdução da perturbação não alterou o bom desempenho do estimador proposto  $mk$ , com relação ao vício e erro médio quadrático, em média.

# Referências Bibliográficas

- [1] AVALOS et al (1993). *Bone Marrow Transplantation* 13, p.133-138.
- [2] AQUINO, E. F. (1969). *Alotransplante Renal Humano: Estudo Comparativo da Sobrevida e das Intercorrências com Dois Esquemas Terapêuticos Imunodepressores*. Tese de Doutorado, Faculdade de Medicina de Ribeirão Preto, USP.
- [3] ABRAMOWITZ & STEGUN (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. NBS AMS, 55, Eds.
- [4] ALTMAN, D. G. (1991). *Practical Statistics for Medical Research*. Chapman and Hall, London, 1<sup>a</sup> ed.
- [5] BENZE, B. G., FRANCO, M. A. P. (1995). *Estimadores Análogos ao de Kaplan-Meier em Modelos Riscos Competitivos Generalizados*. Anais da 5<sup>a</sup> Reunião Regional da Sociedade Brasileira de Matemática Aplicada e Computacional, 186-191, UNESP-Rio Claro.
- [6] BENZE, B. G. (1997). *Análise de Estimadores de Sobrevivência em Estudos de Pequena Escala, sob Riscos Competitivos*. Tese de Doutorado, Universidade de São Paulo.
- [7] BERKSON, J. and GAGE, R. P. (1952). *Survival Curve for Cancer Patients Following Treatment*. Journal of the American Statistic Association, v.47, p.501-515.
- [8] COX, D. R., OAKES, D. V. (1994). *Analysis of Survival Data*. Chapman and Hall, London.
- [9] DAVID, H. A., MOESCHBERGER, M. (1978). *The Theory of Competing Risks*. MacMil, New York.

- [10] EFRON, B. (1979). *Bootstrap Methods: Another Look at the Jackknife*. Annals of Statistics, v.7, p.1-26.
- [11] EFRON, B.; TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York, p.436.
- [12] FLEMING, T. R. (1978). *Nonparametric estimation of Non-Homogeneous Markov Processes in the of Competing Risks*, Annals of Statistics, 6, 1057-1070.
- [13] FLEMING, T. R. (1978). *Asymptotic Distribution Results in Competing Risks Estimation*. Annals of Statistics, v.6, p.1071-1079.
- [14] GAMERMAN, D. (1994). *Bayes Estimation of the Piece-Wise Exponential Distribution*. IEEE Transactions on Reabiliy, v.43, p.128-131.
- [15] KAPLAN, E. L., MEIER, P. (1958). *Nonparametric Estimation from Incomplete Observation*. Journal of the American Statistical Association, v.58, p. 457-481.
- [16] KLEIN, JOHN P.; MOESCHBERGER, MELVIN L., (1997). *Data Set for Survival Analysis*, Second Edition.
- [17] KLEINBAUM, D.G. (1989). *Survival Analysis Module Series*. University of North Carolina at Chapel Hill.
- [18] KIM, J. S. ; PROSCHAN, F. (1991). *Piece-Wise Exponential Estimator of the Survival Function*, v. 40, p. 134-139.
- [19] KITCHIN, J. (1980). *A New Method for Estimating Life Distributions from Incomplete Data*. Doctoral Dissertation, Departament of Statistics, Flórida State University.
- [20] LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons. New York, New York.
- [21] LINK, W. A. (1989). *A Model for Informative Censoring. Theory and Methods*. American Statistical Association. Journal of the American Statistical Association, v. 84, p. 407.

- [22] LITTELL, A. S. (1952). *Estimation of the T-Year Survival Rate from Follow-up Studies Over a Limited Period of Time*. Human Biology, 24, p. 87-116.
- [23] LOUZADA-NETO, F., MAZUCHELI, J., ACHCAR, J.A. (2002). *Análise de sobrevivência e confiabilidade*. Instituto de Matemáticas y Ciencias Afines (IMCA), Pontificia Universidad Católica del Perú. Lima, Peru.
- [24] MILLER, R. G., GONG, G., MUÑOZ, A. (1981). *Survival Analysis*. John Wiley & Sons.
- [25] NELSON, W. (1970). *Hazard Plotting Methods for Analysis of Life Data with Different Failure Models*. J. Qual. Tech., 2, p. 126-149.
- [26] OAKES, D. (1993). *A Note on the Kaplan-Meier Estimator*. American Statistical Association, v.47, p.39.
- [27] PETERSON, A. V. J.(1976). *Bounds for a Joint Distribution Function with Fixed Sub-Distributions Functions. Application to Competing Risks*. Departament of Biostatistics, University of Washington, v. 73, n° 1, p. 11-13.
- [28] R DEVELOPMENT CORE TEAM. (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [29] SALINAS-TORRES, V. H.(1992). *Análise Bayesiana Não Paramétrica na Teoria de Riscos Competitivos*, Tese de Doutorado, IME-USP.
- [30] SALINAS-TORRES, V. H., PEREIRA, C. A., TIWARI, R. C. (1992). *A Nonparametric Bayesian Analysis of Competing Risks Models*. Relatório Técnico, IME-USP.
- [31] SUZARLA, V., VAN RYZIN, J. (1976). *Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observation*. Jornal American Statistic Association, 71, 897-902.
- [32] TSIATIS A. (1991). *A Noidentifiability Aspect of the Problem of Competing Risks*. Proceedings of the National Academy of Science, v. 72, p. 20-22, 1975. KIM, J. S. ;

PROSCHAN, F. *Piece-Wise Exponential Estimator of the Survival Function*, v. 40, p. 134-139.