UNIVERSIDADE DE SÃO PAULO

FACULDADE DE SAÚDE PÚBLICA

DEPARTAMENTO DE EPIDEMIOLOGIA

ANÁLISE DE ESTIMADORES DAS CURVAS DE SOBREVIVÊNCIA
EM ESTUDOS DE PEQUENA ESCALA, SOB DADOS CENSURADOS

Benedito Galvão Benze

Tese apresentada ao Programa de Pós-Graduação em Saúde Pública da Faculdade de Saúde Pública - USP, como parte dos requisitos para a obtenção do título de Doutor em Saúde Pública, área de concentração: Epidemiologia.

Orientador: Prof. Dr. José Maria Pacheco de Souza

Co-orientador: Prof. Dr. Carlos Alberto de Bragança Pereira

São Paulo
outubro de 1997

À minha esposa Maria Júlia

Às minhas filhas Rachel, Débora e Simone

Aos meus pais Valério (in memorian) e Maria Aparecida

Aos meus sogros Júlio e Delfina

Ao meu irmão Alexandre (in memorian)

dedico este trabalho.

AGRADECIMENTOS

Ao meu orientador Prof.Dr. José Maria Pacheco de Souza,
pelo crédito, apoio e orientação.

Ao meu co-orientador Prof.Dr. Carlos Alberto de Bragança Pereira,
pelas, sugestões, incentivo e valiosas contribuições.

Aos Professores Doutores: Maria Aparecida de Paiva Franco, Jorge Oi
e Lael Almeida de Oliveira, pela leitura do trabalho e sugestões

Aos Professores e Funcionários do Departamento de Estatística da
Universidade Federal de São Carlos (UFSCar), pelo apoio e compreen
durante o período de elaboração deste trabalho.

Aos Professores e Funcionários da área de Estatística do Depart
mento de Epidemiologia da Faculdade de Saúde Pública-USP,
pela atenção e carinho que sempre me dispensaram.

Aos alunos (da Estatística-UFSCar): Adrian Roberto Barbosa,
pela elaboração e implementação dos programas computacionais deci
sivos e Eduardo Alves de Amorim, pelos projetos iniciais.

e

Calebe de Paula Bianchini (Ciências de Computação-UFSCar) e Saulo
Portela Parreira (Engenharia Mecânica-USP-São Carlos),
pela valiosa assessoria na solução dos problemas computacionais.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico
(CNPq)

e

Coordenadoria de Apoio à Pesquisa do Ensino Superior
(CAPES),
pelo auxílio financeiro.

RESUMO:

Mediante o estudo do viés e erro médio quadrático, foi comparado o desempenho dos estimadores F-N-P^{*} de Kaplan e Meier (1958) e de Kitchin (1980) e do estimador bayesiano^{**} de Salinas e Pereira (1992), das curvas de sobrevivência sob dados censurados.

Além disso, foi pesquisado e comparado outro estimador F-N-P para esse mesmo fim, que foi chamado estimador modificado de Kitchin (pela mudança realizada na taxa de risco acumulada do estimador de Kitchin nos subintervalos formados pelos tempos consecutivos das ocorrências dos eventos de interesse).

As estimativas são calculadas e comparadas primeiramente em um exemplo de dados clínicos reais de transplantes renais humanos e depois em amostras geradas por simulação a partir de modelos teóricos, assumindo distribuições exponencial e de Weibull.

As simulações indicaram que o estimador de Kaplan e Meier é melhor que os demais, isto é, tem menor erro quadrático médio em todas as circunstâncias abordadas. Neste mesmo sentido o estimador modificado apresentou-se melhor que o de Kitchin.

Com uma priori não informativa, o estimador de Salinas e Pereira teve melhor desempenho que o modificado de Kitchin.

A análise simultânea do desempenho e simplicidade operacional aponta para os estimadores de Kaplan e Meier e Modificado de Kitchin, nessa ordem.

* A sigla refere-se à inferência estatística segundo a metodologia desenvolvida por Fisher, Neyman e Pearson, também conhecida como "inferência clássica" (15,37).

** O termo refere-se à inferência estatística desenvolvida segundo a linha filosófica sugerida por um trabalho de Bayes (1763) (15).

SUMMARY:

The performances of F-N-P^{*} estimators of the survival curve for censored data of Kaplan & Meier (1958), Kitchin (1980) and the bayesian estimator^{**} of Salinas & Pereira (1992) are compared using bias and mean squared error.

A F-N-P estimator, referred to as modified Kitchin, is proposed here and compared with the others. It is based on the change of Kitchin cumulative risks ratios for the sub-intervals given by the consecutive times of occurrence of the events of interest.

The estimates are computed and compared for two cases: first for a real data set related to human renal transplants. And second, for simulated data sets developed from theoretical models for exponential and Weibull distributions.

Simulated results indicated that the Kaplan & Meier estimator is better than the others, that is, it presented least mean squared error in all situations discussed in this work. The results also indicated that the modified estimator is better than the Kitchin estimator.

Using a non-informative "a priori" distribution, the Salinas & Pereira estimator presented better performance than the modified Kitchin estimator.

The simultaneous analysis of the performance and the operational simplicity point out the Kaplan & Meier and modified Kitchin estimators, in this order.

* F-N-P refers to the methodology developed by Fisher, Neyman and Pearson, also known as "classical inference" (15,37).

** Bayesian inference refers to the inference whose development were based on the work of Bayes (1763) (15).

ÍNDICE

sumário.....	iv
Summary	v
Lista de Tabelas	ix
Lista de Figuras	xiii
Anexo.....	xvii

INTRODUÇÃO.....	001
Objetivos.....	007

MATERIAL E MÉTODOS

Cap.1 CONCEITOS BÁSICOS, TERMINOLOGIA, SUPOSIÇÕES TEÓRICAS, MÉTODOS DE COMPARAÇÃO NUMÉRICA E RECURSOS COMPUTACIONAIS:	
1.1. Formulação inicial do problema.....	009
1.2. Algumas considerações teóricas importantes no estabe- lecimento e operacionalização dos estimadores.....	013
1.3. Um exemplo com dados reais.....	015
1.4. Definições, notação e adaptações ao exemplo.....	018
1.5. Métodos de comparação do desempenho dos estimadores da função de sobrevivência sob dados censurados.....	024
1.6. Recursos comutacionais.....	026
Cap.2 ALGUNS TÉCNICAS, SEGUNDO A INFERÊNCIA "FISHER-NEYMAN-PEAR- SON" (F-N-P), PARA ESTIMAR A CURVA DE SOBREVIVÊNCIA:	
2.1. O estimador atuarial	027
2.2. O estimador de Kaplan e Meier	032
2.3. O estimador de Nelson	034
2.4. O estimador exponencial por partes de Kitchin	037
2.5. Comparação das estimativas da curva de sobrevivên- cia segundo os estimadores F-N-P.....	041

Cap.3 ALGUNS MÉTODOS NÃO PARAMÉTRICOS , SEGUNDO A INFERÊNCIA BAYESIANA, PARA ESTIMAR A CURVA DE SOBREVIVÊNCIA:

- 3.1. Conceitos básicos sobre a inferência bayesiana.....043
- 3.2. O estimador de Suzarla e Van Ryzin.....046
- 3.3. Cálculo do estimador de Suzarla e Van Ryzin.....048
- 3.4. O estimador de Salinas e Pereira.....050
- 3.5. O cálculo do estimador de Salinas e Pereira.....051

DESENVOLVIMENTO TEÓRICO E RESULTADOS

Cap.4 EXPRESSANDO OS ESTIMADORES (F-N-P) EM TERMOS DE FUNÇÕES EMPÍRICAS:

- 4.1. Algumas aproximações discretas para o estabelecimen
to dos estimadores em termos de funções empíricas....056
- 4.2. Expressando o estimador de Nelson em termos de fun-
ções empíricas.....060
- 4.3. Expressando o estimador de Kaplan-Meier em termos
de funções empíricas.....061
- 4.4. Expressando o estimador atuarial em termos de fun-
ções empíricas.....063
- 4.5. Expressando o estimador de Kitchin em termos de fun-
ções empíricas.....064

Cap.5 COMPARAÇÃO ENTRE OS ESTIMADORES DAS TAXAS $re(I_m)$:

- 5.1. Interpretação geométrica das estimativas de $re(t)$065
- 5.2. Estimativas da taxa $re(I_m)$067
- 5.3. Comparação das estimativas das taxas $re(I_m)$, conside
rando os estimadores F-N-P apresentados e interpreta
ção geométrica:
 - 5.3.1. Comparação entre as taxas (en) e (ek)068
 - 5.3.2. Interpretação da taxa (km) em termos de á-
reas.....069
 - 5.3.3. Interpretação da taxa (ea) em termos de á-
reas.....070
 - 5.3.4. Comparação entre as taxas (km) e (ea)071

5.3.5. Comparação entre as taxas (ek), (en) e (km)...	072
5.3.6. Comparação entre as taxas (en), (km).....	072
5.3.7. Comparação entre as taxas (ep), (ea) e (km)...	074
5.3.8. Comparação entre as taxas (ep) e (ek).....	074
5.3.9. Definição de elementos de área auxiliares para comparar as taxas (ek) e (en), em relação à taxa (ea).....	076
5.3.10. Comparação entre as taxas (ek) e (ea).....	077
5.3.11. Comparação entre as taxas (en) e (ea).....	078
5.4. Expressão da parte discreta do estimador de Kitchin..	079

Cap 6. PROPOSTA DE UM NOVO ESTIMADOR F-N-P ALTERNATIVO AOS ESTIMADORES F-N-P ESTUDADOS:

6.1. Comparação das formas funcionais na parte discreta dos estimadores F-N-P estudados.....	084
6.2. Proposta do novo estimador F-N-P, não paramétrico, alternativo aos apresentados.....	086
6.3. Uma análise dos fatores que diferenciam os estimadores com a forma funcional da classe 1.....	090
6.4. Cálculo do estimador (mk), para os dados de transplante de rim, doador vivo, esquema A.....	091

Cap 7. SIMULAÇÃO DOS DADOS E COMPARAÇÃO DO DESEMPENHO DOS ESTIMADORES.....

7.1. Esquema de geração das amostras.....	094
7.2. Cálculo do estimador de Salinas-Pereira usando os dados simulados.....	096
7.3. Cálculo do desempenho dos estimadores usando os dados simulados.....	096
7.4. Resultados observados nos dados simulados.....	098
7.5. Considerações finais.....	131

CONCLUSÕES132

REFERÊNCIAS BIBLIOGRÁFICAS134

ANEXO.....140

LISTA DE TABELAS:

Tab. 2.1	Estimativas (eb) , (ea_2) , (km) , (en) e (ek) da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A	042
Tab. 3.1	Estimativas da curva de sobrevivência para os dados de transplante de rim, segundo os estimadores (km) , (ek) , (sv) [para $a(t)=1-0,0005t$] e sp [para $a_1(t)=1-0,0005t$ e $a_2(t)=1-0,0005t$].	055
Tab. 6.1	Estimativas (ek) e (mk) da curva de sobrevivência e da média e mediana, para os dados de transplante de rim, doador vivo, esquema A.	093
Tab. A1	Entradas e saídas, durante a observação de uma amostra de pacientes de transplante de rim, na Unidade de Transplantes Renais do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, entre 14/02/72 a 31/12/78 (doador vivo e esquema terapêutico A) e duração dos intervalos de seguimento.	140
Tab. A2	Contagem do tempo (em dias) de seguimento dos pacientes de transplantes renais da tabela 1A.	141
Tab. A3	Cálculo das estimativas das curvas de subsobrevivência e de sobrevivência "bruta" e da "média bruta", para os dados do exemplo de transplante de rim, doador vivo, esquema A, na UTR do Hospital das Clínicas de Ribeirão Preto.	142
Tab. A4	Estimativa atuarial (ea_1) , da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	143

Tab. A5	Estimativa atuarial (ea2), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	144
Tab. A6	Estimativa de Kaplan-Meier (km), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	145
Tab. A7	Estimativa de Nelson (en), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	146
Tab. A8	Estimativa de Kitchin (ek), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	147
Tab. A9	Estimativas (sv) da curva de sobrevivência e da correspondente média para os dados de transplante de rim, com doador vivo e tratamento A, para $a(t)=1-0,0005t$. ..	148
Tab. A10	Estimativas (sp) da curva de sobrevivência e da correspondente média para os dados de transplante de rim, com doador vivo e tratamento A, para uma priori não informativa uniforme para $\rho=P[Te<To]$	149
Tab. A11	Cálculo da estimativa (mk), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	150
Tab. 7.1	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.0015);\exp(\lambda=0.0015);\pi=0]$	099
Tab. 7.2	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=10;\exp(\lambda=0.0015);\exp(\lambda=0.0015);\pi=0]$	101

Tab. 7.3	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.0015);\exp(\lambda=0.0015);\pi=0.40]$	102
Tab. 7.4	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=10;\exp(\lambda=0.0015);\exp(\lambda=0.0015);\pi=0.40]$	105
Tab. 7.5	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.007);\exp(\lambda=0.03);\pi=0.20]$	107
Tab. 7.6	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=10;\exp(\lambda=0.007);\exp(\lambda=0.03);\pi=0.20]$	109
Tab. 7.7	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.007);\exp(\lambda=0.03);\pi=0.40]$	111
Tab. 7.8	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=10;\exp(\lambda=0.007);\exp(\lambda=0.03);\pi=0.40]$	113
Tab. 7.9	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.0015);Weib(\alpha=0.0015,\beta=1.3);\pi=0.20]$	115
Tab. 7.10	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=10;\exp(\lambda=0.0015);Weib(\alpha=0.0015,\beta=1.3);\pi=0.20]$	117
Tab. 7.11	Desempenho dos estimadores com os dados simulados na configuração $[N=500;n=50;\exp(\lambda=0.0015);Weib(\alpha=0.0015,\beta=1.3);\pi=0.40]$	119

Tab. 7.12 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp($\lambda=0.0015$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].121

Tab. 7.13 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;Weib($\alpha=0.0015,\beta=1.3$));Weib($\alpha=0.0015,\beta=1.3$); $\pi=0$].123

Tab. 7.14 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;Weib($\alpha=0.0015,\beta=1.3$));Weib($\alpha=0.0015,\beta=1.3$); $\pi=0$].125

Tab. 7.15 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;Weib($\alpha=0.0015,\beta=1.3$));Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].127

Tab. 7.16 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;Weib($\alpha=0.0015,\beta=1.3$));Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].129

LISTA DE FIGURAS:

- Fig.1.1 Representação gráfica do processo de contagem durante um seguimento de pacientes submetidos a um certo tratamento.009
- Fig.1.2 Representação gráfica da curva empírica de sobrevivência durante o seguimento de pacientes submetidos a um certo tratamento.010
- Fig.1.3 Comparação das curvas empíricas, que dão as estimativas da verdadeira curva de sobrevivência de pacientes, considerando a presença e a ausência de censuras.012
- Fig.1.4 Entradas (pela ordem de chegada) e saídas de estudo, duração do tempo de acompanhamento, esquema de censura, de pacientes submetidos a transplante de rim, com doador vivo, esquema A.017
- Fig.1.5 Estimativas das curvas de subsobrevivência e de sobrevivência bruta para os dados de transplante de rim, doador vivo, esquema terapêutico A.023
- Fig.2.1 Representação de um subintervalo I_m entre dois tempos consecutivos de "mortes de interesse" te_{m-1} e te_m , contendo tempos censurantes to_1 e to_2029
- Fig.2.2 Estimativas da curva de sobrevivência (ea) para os pacientes de transplante de rim, considerando dois agrupamentos dos dados em subintervalos.031
- Fig.2.3 Estimativa da curva de sobrevivência (km) para os dados de transplante de rim, doador vivo, esquema A.034

Fig.2.4	Estimativa da curva de sobrevivência (e_n) para os dados de transplante de rim, doador vivo, esquema A.	036
Fig.2.5	Estimação da taxa instantânea de risco r_1 , no m -ésimo subintervalo entre os tempos de duas "mortes" consecutivas.	038
Fig.2.6	Estimativa da curva de sobrevivência (e_k) para os dados de transplante de rim, doador vivo, esquema A.	040
Fig.2.7	Comparação das estimativas (e_a), (k_m), (e_n) e (e_k) da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.	041
Fig.3.1	Estimativas (k_m), (e_k) e (s_v), da curva de sobrevivência para os dados de transplante de rim, com doador vivo e tratamento A, admitindo-se $a(t)=1-0,0005t$	049
Fig.3.2	Representação da partição λ em cada subintervalo I_j	052
Fig.3.3	Estimativas (k_m), (e_k) e (s_p) da curva de sobrevivência para os dados de transplante de rim, com doador vivo e tratamento A, admitindo-se uma priori não informativa uniforme para $\rho=P[T_e < T_o]$ e considerando $U_1=U_2=2000$ dias..	054
Fig.5.1	Representação das áreas consideradas na interpretação da estimação das taxas $re(I_m)$, segundo os estimadores (e_k) e (e_n), em uma situação hipotética.	065
Fig.5.2	Representação das áreas, consideradas na interpretação da estimação das taxas $re(I_m)$, segundo o estimador (k_m).	069
Fig.5.3	Quatro possíveis situações sobre o comportamento do processo $V[.]$, em um subintervalo I_m	075

Fig.5.4	Representações gráficas, para uma situação particular, dos elementos de área ΔA e ΔB	076
Fig.5.5	Distribuição das "censuras" ^D e "mortes" ^E no subintervalo I_m	079
Fig.6.1	Comparação gráfica entre as estimativas (ek) e (mk) , da curva de sobrevivência, para os dados de transplante de rim, doador vivo, esquema A.	092
Fig.7.1	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;\exp(\lambda=0.0015),\exp(\lambda=0.0015);\pi=0]$	100
Fig.7.2	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500; n=10;\exp(\lambda=0.0015),\exp(\lambda=0.0015);\pi=0]$	102
Fig.7.3	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;\exp(\lambda=0.0015),\exp(\lambda=0.0015);\pi=0.40]$	104
Fig.7.4	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;\exp(\lambda=0.0015),\exp(\lambda=0.0015);\pi=0.40]$	106
Fig.7.5	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;\exp(\lambda=0.007),\exp(\lambda=0.03);\pi=0.20]$	108
Fig.7.6	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;\exp(\lambda=0.007),\exp(\lambda=0.03);\pi=0.20]$	110
Fig.7.7	Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;\exp(\lambda=0.007),\exp(\lambda=0.03);\pi=0.40]$	112

- Fig.7.8 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;exp(\lambda=0.007),exp(\lambda=0.03);\pi=0.40]$114
- Fig.7.9 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;exp(\lambda=0.0015),Weib(\alpha=0.0015,\beta=1.3);\pi=0.20]$116
- Fig.7.10 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;exp(\lambda=0.0015),Weib(\alpha=0.0015,\beta=1.3);\pi=0.20]$118
- Fig.7.11 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;exp(\lambda=0.0015),Weib(\alpha=0.0015,\beta=1.3);\pi=0.40]$120
- Fig.7.12 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;exp(\lambda=0.0015),Weib(\alpha=0.0015,\beta=1.3);\pi=0.40]$122
- Fig.7.13 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;Weib(\alpha=0.0015,\beta=1.3),Weib(\alpha=0.0015,\beta=1.3);\pi=0]$124
- Fig.7.14 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=10;Weib(\alpha=0.0015,\beta=1.3),Weib(\alpha=0.0015,\beta=1.3);\pi=0]$126
- Fig.7.15 Comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio e viés, para a configuração $[N=500;n=50;Weib(\alpha=0.0015,\beta=1.3),Weib(\alpha=0.0015,\beta=1.3);\pi=0]$128

1.3); $\pi=0.40$].128

Fig.7.16 Comportamento dos estimadores km, ek, mk e sp, com relação ao erro quadrático médio e viés, para a configuração [N=500;n=10;Weib($\alpha=0.0015,\beta=1.3$),Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].130

ANEXO:

Tabelas A1 a A11, contendo detalhes sobre os cálculos das estimativas das probabilidades de sobrevivência e das médias e medianas, segundo os estimadores estudados e considerando dados reais de um seguimento de pacientes de transplantes renais. ...140

INTRODUÇÃO:

A medida da duração da vida de indivíduos tem sido uma das preocupações do homem desde a antigüidade^{14,29,36}. Assim, surgiram métodos como o das tábuas de vida, que continuam sendo aprimorados até hoje^{19,59,60}, para quantificar a mortalidade ou a vitalidade de populações inteiras ou de somente parte delas.

Tanto em estudos de grande escala, com interesse nos tempos de sobrevivência dos indivíduos em alguma população, como nos estudos de pequena escala, com interesse nos tempos de sobrevida em amostras de algum tipo de paciente, existe a possibilidade da ocorrência de observações incompletas ou perdidas, chamadas, genericamente, de censuradas^{37,42,61}.

Para os estudos de grande escala, a utilização das técnicas da tábua de vida, segundo a metodologia de Chiang^{19,20}, tem sido considerada satisfatória na determinação dos modelos de mortalidade e na avaliação dos efeitos provocados em tais modelos quando são eliminados um (ou mais) risco(s) competitivo(s)^{29,50}.

A crescente solicitação por tomada de decisões rápidas também acontece na área da Saúde Pública, principalmente pela gravidade de muitos de seus problemas. Isso tem realçado e estimulado cada vez mais o emprego de estudos baseados em amostras, citando-se alguns mais diretamente relacionados com o assunto aqui tratado: seguimentos de câncer^{32,44,56,57}, de AIDS^{41,63,69}, de transplantes de órgãos^{7,10,17}, bem como estudos sobre métodos contraceptivos⁵³ ou dados de permanências hospitalares²¹. No entanto, faltam indicações sobre quais métodos de estimação dos "modelos de mortalidade" sejam aconselhados em cada uma dessas e outras possibilidades.

O estimador de Kaplan e Meier³³, das probabilidades de so-

brevivência, relativamente a um determinado risco de interesse, tem sido bastante estudado, difundido e utilizado e têm surgido propostas de modificações^{18,24,27,35,43,45,51,58} ou de generalizações desse estimador, como as de Aalen^{1,2,3,4} ou de Fleming^{25,26}.

Na forma, o estimador de Kaplan e Meier é equivalente ao estimador F-N-P atuarial^{14,23}, também comumente empregado na construção de tábuas de vida. No entanto, quando comparado com o atuarial, mostra-se pouco viciado para grandes amostras¹⁶ e fortemente consistente⁶³.

Uma alternativa a estes estimadores, que teoricamente apresenta vantagens em alguns aspectos^{34,49}, é o estimador PEXE* de Kitchin (1980)³⁴, mais conhecido na área de confiabilidade.

Consultando os principais periódicos especializados em Saúde Pública, Epidemiologia ou Biometria, publicados após 1980, não foram encontrados trabalhos que utilizassem esse procedimento. Uma possível explicação para isso seria a sua maior dificuldade de cálculo, em relação aos estimadores atuarial e de Kaplan e Meier.

A expressão básica do estimador de Kitchin é semelhante à do estimador F-N-P de Nelson^{4,46}, também muito usado em confiabilidade^{46,48} e mesmo em análise de sobrevivência¹⁰.

Uma característica comum aos estimadores: atuarial, de Kaplan e Meier e de Nelson, é que suas estimativas da curva de sobrevivência são funções descontínuas à esquerda, nos pontos de "mortes", assumindo o aspecto de escada, com os patamares descendentes com o aumento do tempo, formados entre cada dois pontos consecutivos de "mortes".

O estimador de Kitchin, além das propriedades assintóticas de convergência fraca e forte³⁴, apresenta estimativas que delineiam curvas contínuas, compostas por "pedaços" de funções exponenciais,

* Da sua denominação em inglês: "piecewise exponential estimator".

ao invés dos simples degraus como nos estimadores anteriores.

Tem sido muito divulgada ultimamente uma categoria de estimadores na linha da inferência bayesiana^{12,31,64,67,70}.

Nessa forma de abordagem, a previsão condicionada à informação anteriormente conhecida é modelada por uma distribuição "a priori" e a informação atual, trazida pelos dados amostrais, é representada pela função de verossimilhança. Usando um procedimento básico, chamado "operação bayesiana", reúne-se essas informações em uma distribuição "a posteriori" que atualiza essa previsão^{15,39,52}. Desta forma, com os estimadores bayesianos, existe a possibilidade da incorporação de informações anteriores nas estimativas.

Quando a distribuição "a priori" representar nenhuma ou muito pouca informação conhecida, ela é chamada de "priori não informativa". O estimador bayesiano derivado de tal situação pode ser comparado com estimadores F-N-P, os quais também não incorporam informações anteriores em suas estimativas.

Se o desempenho de um estimador bayesiano, sob uma priori não informativa, for considerado satisfatório em relação ao de um estimador F-N-P, torna-se razoável pensar que sob outra "priori mais informativa" o seu desempenho melhore e até supere o desempenho deste estimador F-N-P.

Novamente, tratam-se de técnicas ainda pouco utilizadas, no Brasil, principalmente em algumas áreas como a de Saúde Pública, mas uma alternativa importante a ser explorada, na tentativa de obter-se resultados cada vez mais confiáveis nas análises estatísticas de dados de saúde. Também, existem algumas dificuldades para a sua implementação, quer seja pelas diferenças de interpretação em relação ao que se pratica na inferência F-N-P, ou pela maior complexidade da tarefa computacional.

Um dos estimadores bayesianos mais recentes é o proposto por Salinas e Pereira (1992)⁷⁰ que, comparado com o estimador bayesiano de Suzarla e Van-Ryzin (1976)⁶⁴, apresenta, por exemplo, a diferença de também levar em conta os tempos de "censuras", no estabelecimento da priori.

As propriedades assintóticas dos estimadores de Suzarla e Van Ryzin e de Salinas e Pereira foram provadas pelos seus autores⁶⁹
⁷⁰.

As curvas estimadas pelos estimadores de Suzarla e Van Ryzin e de Salinas Pereira também assumem o aspecto de "funções escada", com descontinuidades à esquerda dos pontos de "mortes". Mas, os seus patamares apresentam inclinações com decaimento para o final de cada subintervalo onde eles se situam e que vão se acentuando com o aumento do tempo.

Nas situações com tendência a nenhuma informação à priori, os estimadores de Suzarla e Van-Ryzin e de Salinas e Pereira tendem a versões modificadas do estimador de Kaplan e Meier⁷⁰.

A principal pergunta que se formulava antes do início deste estudo era sobre a possibilidade da realização de alguma nova proposta simplificadora, dentro dos métodos de estimação F-N-P comentados até aqui.

A maior contribuição deste trabalho foi justamente neste sentido, com a elaboração de um estimador F-N-P alternativo, semelhante na forma ao de Kitchin, mas diferente na definição da sua taxa de "mortes". A modificação dessa taxa introduziu maior facilidade de cálculo e proporcionou ao estimador resultante (que foi chamado estimador modificado de Kitchin) um desempenho ligeiramente melhor que o do próprio Kitchin.

Na comparação do desempenho desses estimadores também foram incluídos os estimadores de Kaplan e Meier, e o de Salinas e Pereira. O primeiro, pela sua simplicidade e pelo fato de ser o mais

usado atualmente e o segundo para, dentro de condições bastante específicas, explorar as suas potencialidades e melhor avaliar a sua então aparente dificuldade de operacionalização.

O procedimento utilizado nesta comparação baseou-se no cálculo do viés e erro quadrático médio, com dados simulados segundo modelos teóricos conhecidos.

Para dados que se ajustem às situações aqui testadas, os resultados apontam para os estimadores de Kaplan e Meier, de Salinas e Pereira, modificado de Kitchin e de Kitchin, nessa ordem. Por outro lado, aliando-se simultaneamente o desempenho e a simplicidade operacional, os melhores estimadores são, pela ordem, os de Kaplan e Meier e modificado de Kitchin.

Obviamente, os dados reais dificilmente se ajustam às condições estudadas. Nestes casos, quando muito, pode-se ter apenas uma vaga idéia de que a magnitude real das probabilidades de sobrevivência seja melhor retratada por uma curva que decaia mais rapidamente (menor tempo mediano de sobrevivência) ou menos rapidamente (maior tempo mediano de sobrevivência) do que aconteceria caso se utilizasse um estimador "mais tradicionalmente aceito".

Nestas circunstâncias, não existe uma "receita" que indique qual estimador seja o mais apropriado. Entretanto o material contido neste trabalho pode servir de base para uma reflexão e facilitar a opção por uma alternativa que se julgue mais aplicável aos dados e condições que se disponha no momento.

O que se pode dizer é que o estimador modificado é recomendado para as situações em que se deseja uma curva contínua como estimativa da função de sobrevivência. O seu cálculo pode ser realizado manualmente sem muito esforço ou mediante a elaboração de um programa computacional específico.

Neste trabalho foi gerado um programa em linguagem Pascal,

denominado APOIO, para realizar as simulações e calcular o desempenho desses quatro estimadores. Nele existe uma opção de cálculo das estimativas, via entrada de dados, cuja listagem não foi incluída no texto pelo fato de estar inserida, de forma descontínua, na listagem global do programa APOIO que, por sua vez, é muito extensa (mais de duas mil linhas de programação).

Entretanto, o programa APOIO é de domínio público, podendo ser facilmente instalado em micro computadores com processador de no mínimo 100 MHz, memória básica de no mínimo 16 MB e disco rígido de no mínimo 1 GB.

Em síntese, este trabalho reúne e acrescenta conhecimentos julgados úteis para auxiliar o profissional de saúde e de biometria na análise de dados censurados de sobrevivência, desenvolvendo uma metodologia própria de comparação dos principais métodos de estimação.

OBJETIVOS:

O presente trabalho tem os seguintes objetivos:

- Propor e estudar um estimador F-N-P alternativo, aos principais estimadores F-N-P atualmente em uso para estimar a curva de sobrevivência, em relação a um risco específico de "morte", utilizando o relacionamento entre a função de sobrevivência e a função risco acumulado, com o emprego de funções empíricas e mediante a comparação analítica das taxas de risco desses estimadores.
- Comparar o desempenho desse método alternativo com os métodos F-N-P de Kaplan e Meier e de Kitchin e bayesiano de Salinas e Pereira, utilizando simulação, em computador, de dados censurados de sobrevivência.

No capítulo 1, é feita a formulação simples do problema, estabelecendo as condições teóricas básicas e, com a ilustração de um exemplo real, apresenta-se as definições e a notação a serem seguidas. Nos capítulos 2 e 3, descreve-se os estimadores mais diretamente relacionados com o tema e resume-se alguns conceitos básicos de inferência bayesiana, mais citados no texto. Nos capítulos 4 e 5 são feitos os desenvolvimentos teóricos que indicarão os critérios a serem usados, no capítulo 6, para a formulação da proposta de um estimador F-N-P alternativo, para a curva de sobrevivência sob dados censurados. No capítulo 7, é explicado o esquema adotado na simulação dos dados e na comparação dos estimadores, mostrando-se graficamente alguns resultados da simulação. Algumas tabelas são apresentadas nos finais de cada capítulo com o fim de facilitar comparações rápidas. Mas, a maior parte dos registros numéricos importantes para as conclusões foram organizados em tabelas incorporadas no capítulo 7 e final do trabalho, por meio de um Anexo.

MATERIAL E MÉTODOS

1. CONCEITOS BÁSICOS, TERMINOLOGIA, SUPOSIÇÕES TEÓRICAS, MÉTODOS DE COMPARAÇÃO NUMÉRICA E RECURSOS COMPUTACIONAIS.

1.1. Formulação inicial do problema.

Um problema muito comum na área de saúde é o da avaliação do desempenho de um tratamento. Por exemplo, o de uma cirurgia, um medicamento, ou mesmo um conjunto de procedimentos terapêuticos, visando a cura ou ao prolongamento da duração da sobrevivência de pacientes, que tenham semelhantes características e estejam inicialmente no mesmo estágio da doença.

Para simplificar, pode-se imaginar que se tem $n=10$ pacientes nesta situação e que a partir da data de início do tratamento observa-se o tempo decorrido até a recuperação total de cada um deles. Paralelamente, faz-se um acompanhamento contínuo, no decorrer do tempo, verificando-se as quantidades $V(t)$ de pacientes que permanecem sob tratamento, até que todos os n elementos inicialmente considerados no estudo tenham se recuperado. Um exemplo disso é representado graficamente na figura 1.1. (onde estão sendo omitidas as descontinuidades simplesmente por razões estéticas).

Figura 1.1 Representação gráfica do processo de contagem durante o seguimento de pacientes submetidos a um certo tratamento.

Como, no início do estudo tem-se dez pacientes, então, simbolicamente $V[0]=10$. Dois pacientes são curados até o tempo t_2 , sendo que o primeiro evento foi registrado no tempo t_1 e, o segundo no tempo t_2 . Portanto, imediatamente após t_2 , são contabilizados $10-2=8$ pacientes ainda em recuperação, o que é representado por $V[t_2]=8$. No tempo x , entre os tempos t_4 e t_5 , permanecem seis indivíduos em observação, de modo que $V[x]=6$. Em $y=t_6=t_7=t_8$, saem juntos três pacientes, restando dois após esse tempo, isto é, $V[y]=2$.

Para representar que só se dispõe de um paciente após o tempo imediatamente anterior ao tempo t_{10} , pode-se usar a notação $V[t_{10}^-]=1$. A partir do tempo t_{10} recuperam-se todos os pacientes, permitindo escrever que $V[t]=0$, para $t>t_{10}$. Assim, tem-se um processo de contagem $V[t]$, do número de remanescentes ao tempo t , cuja trajetória assume, com o passar do tempo, a forma de escada descendente.

Para facilitar a comparação desses resultados com os fornecidos por qualquer outro grupo de pacientes em situação semelhante, submetidos a esse ou a outro tipo de tratamento, deve-se refazer o

Figura 1.2 Representação gráfica da curva empírica de sobrevivência durante seguimento de pacientes submetidos a um certo tratamento.

gráfico esboçado na figura (1.1). Substitui-se as contagens pelas proporções $p[t]$ de pacientes em recuperação após o tempo t , obtendo-se a curva representada na figura (1.2) acima.

No início, a proporção de sobreviventes é $p[0] = V[0]/10 = 1$. Após o tempo t_2 tem-se $p[t_2] = V[t_2]/10 = 0,8$ e, após x e y , respectivamente, 0,6 e 0,2. Após um instante imediatamente anterior ao tempo t_{10} , tem-se $p[t_{10}^-] = 0,1$.

O tempo médio e mediano de recuperação (ou de sobrevivência) podem ser determinados, respectivamente, por

$$\bar{t} = \frac{1}{10} \sum_{i=1}^{10} t_i = \text{Área sob a curva empírica de sobrevivência}$$

e

$$\hat{M}^d = t, \text{ tal que } p(t) = 0,5.$$

A curva empírica de sobrevivência é uma estimativa da verdadeira curva de sobrevivência que, teoricamente, espera-se para qualquer outro grupo de pacientes em condições similares às do grupo em questão. Além da rápida avaliação das estimativas da média e mediana, ela permite observar outros aspectos relevantes a serem considerados no posterior estabelecimento dos modelos de previsão. Por exemplo, torna-se importante observar o modo como ocorrem os eventos para se conjecturar as possíveis consequências sobre a forma da verdadeira distribuição do tempo de sobrevivência.

Quando todas as saídas acontecem pela ocorrência comprovada do evento de interesse, como a atribuição da cura da doença ao tratamento sob estudo (ou a morte do paciente, tendo a doença como a causa principal), a adoção da curva empírica de sobrevivência constitui um procedimento razoável de estimação, pois satisfaz algumas propriedades desejáveis requeridas pela inferência estatística.

Entretanto, um fator complicador que frequentemente ocorre em situações desse tipo é a possibilidade da saída dos indivíduos

por causas diferentes daquela escolhida para estudo. É o que se denomina "censura", gerando os chamados "dados censurados". Nestes casos, dispõe-se somente dos registros até um certo instante que, ou correspondem à data do último retorno de um paciente, com o qual se perdeu o contato, ou ao momento em que um paciente foi subtraído do seguimento, após a ocorrência de um evento estranho à pesquisa.

Nestas situações, o emprego da curva empírica dará uma estimativa "bruta" da verdadeira curva de sobrevivência, inclusive, produzindo subestimativas em relação ao que realmente ocorreu, pois os valores censurados t^* são menores que os verdadeiros tempos de duração na amostra. Por exemplo, poder-se-ia ter 30% de censuras, representadas pelas atribuições de durações: t_2^* (menor que a verdadeira duração t_2), t_6^* (menor que a verdadeira duração t_6) e t_9^* (menor que a verdadeira duração t_9). Em uma situação real esses tempos t_2 , t_6 e t_9 não seriam conhecidos pelo pesquisador. Na figura 1.3 compara-se as estimativas da curva de sobrevivência considerando ou não a ocorrência destas censuras.

Figura 1.3 Comparação das curvas empíricas, que dão as estimativas da verdadeira curva de sobrevivência de pacientes, considerando a presença e a ausência de censuras.

Assim, conclui-se que só a utilização deste procedimento para a avaliação do desempenho de tratamentos poderá, na maioria das vezes, aumentar o risco de se cometer sérios e irreparáveis enganos.

Deve-se então procurar alternativas que ao levarem em conta o efeito dessas censuras, diminuam essa possibilidade.

Existem vários métodos que podem ser usados para a construção da curva de sobrevivência que levem em consideração as censuras. Os que não usam suposições quanto à forma da distribuição do tempo de duração dos pacientes são chamados de métodos não paramétricos. Como estes têm sido tradicionalmente os mais explorados na área de saúde, optou-se por este tipo de abordagem. A questão é de decidir-se qual deles seja o mais apropriado para os dados de que se disponha.

1.2. Algumas considerações teóricas importantes no estabelecimento e operacionalização dos estimadores.

Nas aplicações reais, existem as possibilidades de, nas distribuições do tempo de duração das observações, ocorrerem:

- 1- pontos de salto em comum, (isto é, a ocorrência de empates nos tempos de duração) envolvendo distintas causas de saída de seguimento, às quais ficam sujeitos os indivíduos da amostra e/ou
- 2- pontos de empates, dentro de uma mesma causa.

As dificuldades teóricas, provenientes do primeiro desses problemas, no estabelecimento dos estimadores ou estudo de suas propriedades, têm sido contornadas pela adoção em diferentes graus de exigência, dependendo do autor, de "suposições simplificadoras" sobre a continuidade das funções envolvidas. Por exemplo, variam desde as suposições mais restritivas como em Breslow e Crowley (1974)¹⁶, Pe-

terson (1977)⁵⁵, Suzarla e Van Ryzin (1976)⁶⁴, Salinas e Pereira (1992)⁷⁰, até as suposições mais gerais como em Tsai (1986)⁶⁸, onde teoricamente não são impostas restrições quanto às ocorrências de pontos de salto em comum.

Do ponto de vista prático, as possíveis dificuldades computacionais, geradas por esse tipo de repetição no conjunto de dados, podem ser contornadas, adaptando-se às sugestões dadas em Kaplan e Meyer (1958)³³ e Berliner e Hill (1988)¹². Isto é, pode-se considerar inicialmente que os pontos de "censuras" (referentes às durações de tempo até o acontecimento de um outro evento fora do evento de interesse), ocorram ligeiramente depois dos pontos de "mortes" (referentes às durações de tempo até o acontecimento do evento de interesse). Assim, esses empates podem ser desfeitos, mantendo-se intactos os tempos de "morte" e adicionando aos tempos de censura que lhes sejam coincidentes, pequenas frações de tempo (em relação à unidade de medição utilizada), até não mais existir este tipo de justaposição.

No segundo problema, tem-se os tipos de empates que podem persistir dentro de uma mesma causa, depois de realizada a adaptação descrita acima. Por exemplo, podem continuar ocorrendo empates envolvendo somente os tempos de "mortes" e/ou somente os tempos de "censuras". Neste caso, pode-se adotar um procedimento semelhante ao utilizado no problema 1, quando a "eliminação" destas segundas justaposições, no conjunto de dados, se fizerem necessárias para permitir o emprego de um determinado estimador ou versão dele.

Neste trabalho são consideradas censuras aleatórias que podem ser provocadas por diferentes causas como: chegada dos pacientes de forma aleatória, conforme a data do diagnóstico, e truncamento à direita das durações ainda em acompanhamento no momento da avaliação do tratamento e/ou atribuição das "saídas" a motivos diferentes daquele de interesse principal no estudo (como, por exemplo, as de-

sistências na continuação ao tratamento). Também, é considerado o esquema de censura "não informativo", isto é, não se supõe outra informação nos dados censurados da amostra, além da duração no tempo de seguimento e a causa de finalização de observação. Assim, o "mecanismo de produção" das censuras não será explicitamente considerado nos métodos de estimação das probabilidades de sobrevivência.

Além disso, admite-se a suposição de independência entre os tempos potenciais de "mortes"^o e os de "censuras"^o, o que pode de 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 estudo⁴⁰.

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 não são simultâneos, os tempos censurados também podem ser diferentes, ou seja, são progressivamente censurados³⁷ e satisfazem à definição de censuras à direita, do tipo I (conforme é estabelecido na teoria de análise de sobrevivência)⁴².

Os demais conceitos e terminologia da teoria de análise de sobrevivência, bem como os principais métodos não paramétricos F-N-P e bayesianos de estimação da curva de sobrevivência, são apresentados, adaptados e ilustrados, nos próximos capítulos, por meio do seguinte exemplo com dados reais, envolvendo o tipo de censuras descrito anteriormente.

1.3. Um exemplo com dados reais.

Consideremos o conjunto de dados de Aquino (1979)^o, ex-

traído de Oliveira (1981)⁴⁸, que se referem a 25 (vinte e cinco) 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 (esquema A), 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 pelo truncamento à direita das observações. 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.

Como indicam os registros (tabela A1, do Anexo), os primeiros transplantes, nestes pacientes, só foram realizados a partir do início de 1972. Assim, foi considerado, como período de acompanhamento, o intervalo de 14/02/72 a 31/12/78, o que é ilustrado no esquema abaixo.

O símbolo # está assinalando os anos bissextos.

ORDEM DE ENTRADA

Figura 1.4 Entradas (pela ordem de chegada) e saídas de estudo, duração do tempo de acompanhamento, esquema de censura, de pacientes submetidos a transplante de rim, com doador vivo, esquema A.

Na figura (1.4) representa-se, dentro do período de acompanhamento, as entradas (isto é, o tempo correspondente à cirurgia), as saídas de estudo e a correspondente duração (em dias) do tempo de seguimento, para cada paciente na amostra estudada, seguindo a ordem em que aconteceram as entradas para o tratamento. Note-se que a duração média de seguimento foi de 1008,6 dias e a mediana foi de 1272 dias (daqui em diante chamadas, respectivamente, duração média e mediana "brutas").

A duração de tempo entre as entradas e saídas 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. Observa-se que 48 % da amostra é constituída de observações censuradas, devido ao truncamento à direita provocado pela decisão de avaliar o resultado do tratamento até o 2512^o dia de seguimento dos pacientes.

1.4. Definições, notação e adaptações ao exemplo.

Competindo contra a "sobrevivência" dos indivíduos, isto é, contra a sua permanência em observação, imagina-se um conjunto $R = \{R_1, R_2, \dots, R_k, \dots, R_K\}$, de riscos²², conduzindo a um conjunto $C = \{C_1, C_2, \dots, C_k, \dots, C_K\}$, de causas de saídas de seguimento; estes riscos são chamados de riscos competitivos. Associado aos conjuntos R e C tem-se um mesmo conjunto J de índices, que é formado pela união dos subconjuntos de índices E (indicando os riscos relacionados aos eventos de interesse principal) e O (indicando os riscos relacionados aos outros eventos não de interesse), isto é $J = E \cup O$, de modo que $E \cap O = \emptyset$.

Assim, os elementos dos subconjuntos E e O podem distinguir os riscos competitivos [ou causas], respectivamente, de "mortes"^E, pertencentes a um conjunto de riscos RE [ou de causas CE] de interesse, e de "censuras"^O, pertencentes a um conjunto de riscos RO [ou de causas CO] não de interesse. Desse modo, tem-se $R = RE \cup RO$ e $C = CE \cup CO$, conforme é esquematizado no quadro a seguir.

No exemplo apresentado adota-se a convenção:

$E = \{1\}$, referente ao óbito por problemas no enxerto,

$O = \{2\}$, referente ao truncamento do tempo de sobrevida do enxerto.

Portanto:

$$J = \{1, 2\}.$$

$$RE = \{R1\}, \quad RO = \{R2\} \text{ e } CE = \{C1\}, \quad CO = \{C2\}.$$

$$\text{e } R = \{R1, R2\} \text{ e } C = \{C1, C2\}.$$

Como, neste exemplo, o conjunto de riscos de interesse RE e o correspondente conjunto de causas de interesse CE são formados, cada qual, por um único elemento, pode-se simplificar a notação, fazendo: $RE \equiv R1$ e $CE \equiv C1$. O mesmo se aplica a RO e CO , fazendo $RO \equiv R2$ e $CO \equiv C2$. Em consequência, $E \equiv 1$ e $O \equiv 2$, respectivamente para $k=1$ e $k=2$.

A variável de maior interesse é o tempo potencial de sobrevida do enxerto se somente um RE fixado estivesse agindo contra seu funcionamento, sendo indicada por TE . Quando RE é composto de somente um risco Re (com o índice e indicando um risco específico) então $TE \equiv Te$. O tempo Te passa a ser a duração que teoricamente ocorreria no acompanhamento de um indivíduo, nas condições em estudo, caso fosse possível observá-lo sob a influência isolada do risco Re , até o aparecimento da correspondente causa Ce da sua finalização de seguimento.

Deseja-se estimar a curva da função de sobrevivência "líquida" $Se(t)$, definida pelas probabilidades de que o tempo potencial de sobrevivência Te exceda a um tempo positivo pré-estabelecido t , isto é:

$$Se(t) = P[Te > t], \quad 0 < t < \infty. \quad (1.1)$$

Também tem-se interesse em outras funções que estão associadas a $Se(t)$, como as que medem as intensidades (ou "forças") acumulada e instantâneas de "morte" devidas ao risco Re e conhecidas, respectivamente, como função risco acumulado $Re(t)$ e a taxa instantânea de risco $re(t)$.

Estas funções são definidas, respectivamente, por:

$$Re(t) = - \ln \{Se(t)\} \quad 0 < t < \infty \quad e \quad (1.2)$$

$$re(t) = \frac{d}{dt} [Re(t)] = \frac{fe(t)}{Se(t)}, \quad 0 < t < \infty, \quad (1.3)$$

onde: $\ln(\cdot)$ é o logaritmo neperiano da função e $fe(\cdot)$ é a função densidade de probabilidade da variável aleatória Te .

Por (1.3), o termo $re(t)dt$ pode ser interpretado como a probabilidade condicional de que um indivíduo, tendo "sobrevivido à idade t , "morra" no intervalo de tempo $(t, t+dt)$ ¹⁴.

Mas, o que realmente é observado, são os n ($=25$, no exemplo) tempos T_i de duração no seguimento, correspondentes a cada indivíduo i da amostra estudada que permaneceu exposto a todos os K ($=2$, no exemplo) riscos, até a sua saída de observação. Essa saída é atribuída à ocorrência de uma única causa Ce , do conjunto considerado de causas, a qual é associada somente a um único risco Re , do conjunto de todos K riscos então previstos.

Supõe-se assim, que o tempo observado T_i , de duração do i -ésimo indivíduo da amostra, seja o menor dos k - i -ésimos tempos potenciais de duração T_{k_i} , ($k=1,2,\dots,K$; $i=1,2,\dots,n$) atribuíveis a esse indivíduo quando se leva em conta isoladamente cada risco a que ele está exposto.

$$T_i = \min_{1 \leq k \leq K} \{ T_{k_i} \}, \quad i=1,2,\dots,n. \quad (1.4)$$

No exemplo, tem-se:

$$T_i = \min \{ T_{1_i}, T_{2_i} \}, \quad i=1,2,\dots,25.$$

Fixando um risco Re de interesse, observa-se os n tempos T_i na amostra (algumas vezes, representados simplificada-mente por T) e as correspondentes causas Ce_i de saídas de segmento (às vezes simplifi-cadas para Ce). Então, pode-se estimar as chamadas funções de subsobrevivência $Se_T(t)$ e função de sobrevivência bruta $S_T(t)$. Depen-dendo dos métodos utilizados^{55,71}, as estimativas dessas funções po-dem conduzir indiretamente às estimativas das funções de sobrevivên-cia e funções associadas.

As funções de subsobrevivência são definidas pela probabi-lidade de que as durações T_i (aqui simplesmente representados por T) excedam a um tempo t positivo pré-estabelecido e que sejam finaliza-das pela causa Ce fixada, isto é:

$$Se_T(t) = P[T > t, Ce], \quad 0 < t < \infty. \quad (1.5)$$

Pode-se estimá-las por meio das funções empíricas de subso-brevivência definidas por:

$$Se_T^n(t) = \frac{1}{n} \sum_{i=1}^n I[T_i > t; Ce_i], \quad (1.6)$$

com a função indicadora

$$I[.] = \begin{cases} 1, & \text{se } T_i > t \text{ e } T_i = Te_i \\ 0, & \text{se } T_i > t \text{ e } T_i \neq Te_i \end{cases},$$

dentro do somatório e a letra n indicando o "estimador empírico".

Note-se que se está estimando as probabilidades de sobrevivência ao tempo t , com interesse em um risco específico R_e , mas com os n indivíduos da amostra sujeitos à ação simultânea deste e de todos os outros riscos concorrendo contra as suas "vidas".

A função de sobrevivência bruta é definida pela probabilidade de que as durações no seguimento excedam aos tempos pré-fixados $t \geq 0$, sem especificar um particular risco de interesse, dentre todos aqueles competindo para a eliminação dos indivíduos. Isto é:

$$S_T(t) = P[T > t], \quad 0 < t < \infty. \quad (1.7)$$

Uma estimativa dessa função de sobrevivência bruta é dada pela sobrevivência bruta empírica:

$$S_T^n(t) = \frac{1}{n} \sum_{i=1}^n I[T_i > t], \quad (1.8)$$

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

Essas estimativas foram calculadas para o exemplo aqui tratado¹¹. Os resultados estão organizados na tabela A3, do Anexo e representados na figura 1.5 a seguir.

Observa-se que as curvas estimadas são monótonas decrescentes, descontínuas (à esquerda), nos pontos onde ocorrem os eventos, formando patamares nos subintervalos entre cada dois pontos de "saltos" 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.

Figura 1.5 Estimativas das curvas de subobrevivência e de sobrevivência bruta para os dados de transplante de rim, doador vivo, esquema terapêutico A.

Nota-se, pelos cálculos ilustrados na tabela A3, do Anexo que a área sob a estimativa da curva de sobrevivência bruta tem o mesmo valor da estimativa da média bruta, citada anteriormente. A estimativa da correspondente duração mediana (aqui determinada por interpolação linear, no subintervalo em que a estimativa da função de sobrevivência assume o valor 0,5) é dada por 1249,5 dias (valor a ser utilizado daqui para a frente nas comparações).

A função de sobrevivência bivariada é definida por:

$$S(t_e, t_o) = P[T_e > t_e; T_o > t_o], \quad 0 < t_e < \infty \quad \text{e} \quad 0 < t_o < \infty, \quad (1.9)$$

representando a probabilidade dos tempos potenciais de sobrevivência, relativamente aos riscos R_e e R_o , respectivamente, excederem os valores positivos t_e e t_o .

Duas propriedades são particularmente úteis. A primeira, (por exemplo, Peterson⁵⁵) estabelece, para $t_e=t_o=t$, que o relacionamento entre essa função e a função de sobrevivência bruta é o seguinte:

$$S(t, t) = S_T(t). \quad (1.10)$$

A segunda é a que relaciona as funções de subsobrevivência com a de sobrevivência bruta (Peterson⁵⁵ ou Salinas⁷⁰, por exemplo):

$$S_T(t) = Se_T(t) + So_T(t). \quad (1.11)$$

Assim, a função empírica de sobrevivência bruta pode ser escrita como:

$$S_T^n(t) = Se_T^n(t) + So_T^n(t). \quad (1.12)$$

1.5. Métodos de comparação do desempenho dos estimadores da função de sobrevivência sob dados censurados.

Seja $Se(.)$ um estimador da função de sobrevivência $Se(.)$.

O viés $EM[.]$ deste estimador, em um tempo t , é definido por:

$$EM[Se(t)] = E[Se(t)] - Se(t), \quad (1.13)$$

onde o símbolo $E[.]$ indica a esperança de uma variável aleatória.

Diz-se que o estimador $Se(.)$ é não viciado se, para todo

tempo $t \geq 0$,

$$E[Se(t)] = 0,$$

isto é, se

$$E[Se(t)] = Se(t). \quad (1.14)$$

A qualidade do estimador $Se(\cdot)$, em termos de sua precisão, é medida pelo erro quadrático médio que, em um tempo $t \geq 0$, é definido por:

$$\begin{aligned} EQM[Se(t)] &= E[(Se(t) - Se(t))^2] \\ &= VAR[Se(t)] + EM^2[Se(t)], \end{aligned} \quad (1.15)$$

onde o símbolo $VAR[\cdot]$ indica a variância de uma variável aleatória e EM^2 indica o viés ao quadrado.

Quanto menor for o erro quadrático médio, isto é, quanto menor for o viés e a variância do estimador, tanto melhor é o seu desempenho.

Neste trabalho, o desempenho dos estimadores foi estudado em configurações do tipo:

$$[N; n; Se; So; \pi], \quad (1.16)$$

onde N é o número de amostras, n é o tamanho de cada amostra, Se e So são os modelos teóricos conhecidos assumidos na simulação computacional, respectivamente, dos tempos de "mortes" e de "censuras" e π é a porcentagem de censuras em cada uma das N amostras da configuração.

Os modelos teóricos assumidos nas simulações foram a distribuição exponencial de parâmetro λ e/ou a distribuição Weibull de parâmetros α e β . Isto é,

$$Sk(t) = \exp(-\lambda t), \quad t \geq 0 \quad (1.17)$$

$$e \quad S_k(t) = \exp(-(\alpha t)^\beta), \quad t \geq 0, \quad (1.18)$$

onde $k=e$ (referindo-se às "mortes^e") ou $k=o$ (referindo-se às "censu^oras") e foram atribuídos valores aos parâmetros λ , α , β .

Em uma dada configuração o viés e o erro quadrático médio, em cada ponto t , foram estimados, respectivamente, por:

$$e \quad \text{EM}(t) = \frac{1}{N} \sum_{j=1}^N [Se(t) - Se(t)]_j, \quad (1.19)$$

$$\text{EQM}(t) = \frac{1}{N} \sum_{j=1}^N [Se(t) - Se(t)]_j^2, \quad (1.19)$$

1.6. Recursos computacionais.

Para realizar as simulações e os cálculos foi elaborado um programa computacional, denominado APOIO, usando linguagem PASCAL (softwares: Turbo Pascal, versão 5.0 e Borland Pascal, versão 7.0).

Para verificar a confiabilidade das seqüências geradas utilizou-se análises gráficas e os testes de aderência Qui-quadrado e Kolmogorov e Smirnov (com os programas estatísticos: Statistica, versão 5.0 e Minitab, versão 10.1).

O texto foi confeccionado com o editor Chi-Writer, versão 3.15, e para a construção dos gráficos utilizou-se os programas Excel, versão 5.0 e Harvard Graphics, versão 10.2.

O equipamento computacional utilizado consistiu de computador PC, com micro-processador "Pentium" de 133 MHz, disco rígido de 1,7 GB e memória básica de 16 MB.

2. ALGUMAS TÉCNICAS, SEGUNDO A INFERÊNCIA " FISCHER-NEYMAN-PEARSON " (F-N-P), PARA ESTIMAR A CURVA DE SOBREVIVÊNCIA.

2.1. O estimador atuarial (ea)^{14,23}.

Esta técnica é muito utilizada, principalmente em estudos epidemiológicos 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 Δ_x pode variar com x) e não existe a possibilidade de se identificar a ordem em que ocorreram cada uma das "mortes^e" e/ou cada uma das "censuras^o" dentro de cada subintervalo.

A expressão geral deste estimador é dada por:

$${}^{ea}Se(t) = \prod_{x \leq t} \left[1 - \frac{d_x}{l_x - (1/2)(u_x + w_x)} \right] = \prod_{x \leq t} \left[1 - \frac{d_x}{l_x - l'_x} \right] = \prod_{x \leq t} [1 - q_x], \quad (2.1)$$

onde:

- l_x é o número de indivíduos em risco no início do subintervalo I_x ;
- d_x é o número de "mortes^e" (isto é, provocadas por uma causa específica C_e de interesse) em I_x ;
- u_x é o número de perdidos de segmento em I_x (incluindo os que "morreram" por outras causas, sem interesse imediato ao estudo e as perdas de contato com os pacientes) e
- w_x é o número de "retirados vivos" de seguimento em I_x (devido ao truncamento à direita, pelo término da pesquisa ou qualquer outra causa, que tenha impedido o prosseguimento das observações, quando na realidade se saiba que os indivíduos tenham continuado "vivos").

Note-se que:

$u_x + w_x$ é o número de "censuras^o" ;

q_x estima a probabilidade de um indivíduo vir a "morrer (pela causa de interesse)" no subintervalo I_x , dado que "sobreviveu" ao subintervalo $I_{x'}$, $x' < x$ e

ea

Se(t) é um estimador da probabilidade de sobrevivência para todo subintervalo I_t , considerando o indivíduo exposto somente ao risco Re (com ea indicando que o estimador é o atuarial).

Na elaboração deste método considera-se que, em média, as "censuras^o" em cada subintervalo I_x permaneçam em risco durante a metade deste subintervalo, daí a função do termo

$$l'_x = \frac{u_x + w_x}{2}$$

na expressão do estimador q_x .

Para ilustrar o cálculo desse estimador, são construídas tábuas de vida para duas situações, ambas envolvendo os dados de transplantes de rins do exemplo aqui tratado. Na primeira os tempos até as ocorrências dos eventos são transformados em meses (conforme o procedimento exposto na sexta coluna da tabela A1, no Anexo). Em seguida, são agrupados em subintervalos I_x , de comprimento fixo de um mês, variando de (0,1] até (64,65]. As correspondentes estimativas são simbolizadas por (ea1), para indicar que foram calculadas pelo método atuarial e com a utilização do primeiro agrupamento. Os detalhes dos cálculos são apresentados na tabela A4, no Anexo.

Na segunda situação, adota-se o seguinte procedimento: coloca-se em ordem crescente a sequência observada dos tempos de ocorrência dos eventos, isto é:

$$0 < t_1 < t_2 < t_3 < \dots < t_{j-1} < t_j < \dots < t_{(N-1)} < t_N, \quad (2.2)$$

onde o índice "j" [$1 \leq j \leq N \leq n$] é um inteiro positivo indicando a

ordem da observação e cada t_j é o valor observado do tempo T_e de "morte^e" ou do tempo T_o de "censura^o". O índice j pode não atingir o valor correspondente ao tamanho n da amostra devido à possibilidade da ocorrência de empates.

Em seguida, considera-se os subintervalos $I_m = (t_{e_{m-1}}; t_{e_m}]$, determinados por cada dois pontos consecutivos $t_{e_{m-1}}$ e t_{e_m} , de "mortes^e", para $m=1, \dots, M$ (com $t_{e_0}=0$ e $M \leq N$, pelas possibilidades tanto da ocorrência de censuras como de empates nos pontos de "mortes"). A letra minúscula "e" está sendo aqui usada para ressaltar que os pontos são de "mortes" devido a uma causa fixada C_e de interesse. A letra minúscula "m" especifica a ordem dos valores dos tempos que correspondem somente a essas "mortes^e". Para representar o número de "mortes^e" no intervalo $(0, t]$ usa-se a notação $M(0, t]$. A figura 2.1 fornece maiores esclarecimentos sobre essa notação.

Figura 2.1 Representação de um subintervalo I_m entre dois tempos consecutivos de "mortes de interesse" $t_{e_{m-1}}$ e t_{e_m} , contendo dois tempos censurantes t_{o_1} e t_{o_2} .

Por exemplo, lembrando que $e=1$ (para morte) e $o=2$ (para censura), e $n=25$ observações, 13 das quais são mortes e 12 são censuras, houve ocorrência nos seguintes tempos:

$$t_1 = 40, t_2 = 43, t_3 = 89, \dots, t_{10} = 956, t_{11} = 970, \dots, t_{24} = 1713, t_{25} = 1973$$

ou

$$t1_1 = 40, t1_2 = 43, t1_3 = 89, \dots, t2_1 = 956, t2_2 = 970, \dots, t2_{12} = 1713, t1_{13} = 1973.$$

Como não existem empates, tem-se $1 \leq j \leq 25$, isto é, $N = n = 25$ e $M = 13$ coincide com o número de mortes $M(0, 1973]$.

Define-se:

$V[te_{m-1}]$ = número de indivíduos em risco no início do subintervalo I_m , isto é, o número de indivíduos em risco imediatamente após o $(m-1)$ -ésimo tempo observado de "mortes^e";

$V[te_m^-]$ = número de indivíduos em risco no final do subintervalo I_m , isto é, o número de indivíduos em risco imediatamente antes do m -ésimo tempo observado de "mortes^e";

$M[te_m]$ = número de "mortes^e" registradas no subintervalo I_m , isto é, exatamente no ponto te_m ;

$W[I_m]$ = número de "censuras^o" registradas no subintervalo I_m .

Assim, reescreve-se o estimador (2.1) na seguinte forma:

$$Se(t) = \prod_{te_m \leq t} \left[1 - \frac{M[te_m]}{V[te_{m-1}] - \frac{W[I_m]}{2}} \right], \quad (2.3)$$

notando que

$$\frac{M[te_m]}{V[te_{m-1}] - \frac{W[I_m]}{2}} = \frac{M[te_m]}{V[te_{m-1}] - W[I_m] + \frac{W[I_m]}{2}} = \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}}. \quad (2.4)$$

O primeiro membro da expressão (2.4), é o estimador correspondente ao estimador atuarial q_x na expressão (2.1)³⁸. Se $W[I_m]=0$, o estimador (2.3) se reduz à função de sobrevivência bruta empírica definida por (1.9).

Os detalhes dos cálculos das estimativas segundo este procedimento são apresentados na tabela A5 do Anexo. Como mostra a figura 2.2, as configurações das estimativas da curva de sobrevivência podem apresentar diferenças, conforme o tipo de agrupamento empregado.

tiva da curva de sobrevivência seria a de curvas decrescentes, a partir do ponto (0,1), "suavizadas" por segmentos de reta, ligando os extremos inferiores dos patamares²⁸.

2.2. O estimador de Kaplan e Meier (km)³³.

Uma técnica que tem sido muito divulgada e utilizada na área de saúde é a do estimador não paramétrico de Kaplan-Meier. Para operacionalizá-lo deve-se primeiro fazer a ordenação (2.2), dos tempos de ocorrência dos eventos.

A sua expressão geral, escrita na notação definida em (2.3), é dada por:

$$\text{km} \quad \text{Se}(t) = \begin{cases} 1, & M(0,t)=0 \text{ ou } t=0 \\ \left[1 - \frac{M[te_m^-]}{V[te_m^-]} \right], & M(0,t) \geq 1 \text{ e } t \leq t_N \\ 0, & t_N = te_M \text{ e } t \geq t_N \\ \text{indefinido,} & t_N \neq te_M \text{ e } t > t_N \end{cases} \quad (2.5)$$

Conforme observa-se na primeira linha da expressão (2.5), a estimativa da função de sobrevivência no tempo $t=0$ assume o valor 1. Na terceira linha vê-se que, se o último tempo de ocorrência corresponder a uma "morte" de interesse, a estimativa da curva atinge o valor zero, a partir deste ponto. Caso contrário, ela permanece até o ponto com o mesmo valor já assumido na última ocorrência de "morte" e é assumida indefinida depois deste ponto.

A expressão básica do estimador é dada na segunda linha e apresenta uma forma semelhante à da expressão básica do estimador atuarial (2.3). Os termos do produtório são da forma

$$\frac{V[te_m^-] - M[te_m]}{V[te_m^-]} = \frac{[No. \text{ em risco em } te_m^-] - [No. \text{ de "mortes" em } te_m]}{[No. \text{ em risco em } te_m^-]} \quad (2.6)$$

Para cada tempo t fixado, no intervalo $(0, t_N]$, faz-se o produto acumulado destes termos, calculados em todos os pontos te_m de ocorrência de "mortes"^e 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 $V[te_m^-]$.

Uma das vantagens desse estimador, em relação ao atuarial, é ser pouco viciado para amostras grandes¹⁶. Além disso, o estimador de Kaplan e Meier apresenta consistência forte⁶³.

Quando não se tem mais de uma "morte"^e no mesmo ponto (isto é, quando não existem empates com tais ocorrências), como está acontecendo no exemplo, os termos representados por (2.6) assumem a forma

$$\frac{[n - j + 1] - 1}{[n - j + 1]} = \frac{n - j}{n - j + 1}, \quad (2.7)$$

onde n é o tamanho da amostra e são considerados somente os índices j 's correspondentes aos postos dos tempos de "mortes"^e.

Os cálculos das estimativas das probabilidades de sobrevivência, pelo estimador de Kaplan e Meier, para os dados de transplantes de rim, são apresentados na tabela A6, no Anexo. Obtem-se, como resultado, uma curva descontínua nos pontos de "mortes"^e, decrescendo com o passar do tempo e assumindo a forma de patamares, nos subintervalos formados entre dois tempos consecutivos de ocorrências.

A estimativa da média é de 1246,4 dias (23,6% maior que a duração média "bruta" e 1 % maior que a estimativa ea2. A estimativa da mediana é de aproximadamente 1522,4 dias (21,8% maior que a da me

diana bruta e 2,5% maior que a da estimativa (ea2) da mediana.

A estimativa da correspondente função risco acumulado pode ser obtida, por (1.2), calculando-se

$$R1(t) = -\ln(S1(t)) \tag{2.8}$$

somente nos pontos t onde ocorrem as "mortes"^{e=1}.

Figura 2.3 Estimativa da curva de sobrevivência (km) para os dados de transplante de rim, doador vivo, esquema A.

2.3. O estimador de Nelson (en)^{4,46}.

Este estimador é muito empregado em experimentos indus-

triais, no estudo da confiabilidade de peças e equipamentos. Esta técnica não tem sido encontrada em aplicações na área de saúde pública. As suas estimativas da curva de sobrevivência são obtidas pela aplicação da expressão (1.2), após o cálculo das estimativas da função risco acumulado mediante a utilização do estimador de Nelson. As suas propriedades assintóticas foram também estudadas por Aalen⁴, do ponto de vista de processos de contagem.

O estimador de Nelson para a função risco acumulado é definido por:

$$\text{en} \quad \text{Re}(t) = \sum_{t_{e_m} \leq t} \frac{M[t_{e_m}]}{V[t_{e_m}]} \quad (2.9)$$

e em consequência o estimador da função de sobrevivência é dado por:

$$\text{en} \quad \text{Se}(t) = \begin{cases} 1, & M(0,t)=0 \text{ ou } t=0 \\ \exp \left[- \sum_{t_{e_m} \leq t} \frac{M[t_{e_m}]}{V[t_{e_m}]} \right], & M(0,t) \geq 1 \text{ e } t \leq t_{e_M} \\ \text{indefinido}, & t_N \neq t_{e_M} \text{ e } t > t_N. \end{cases} \quad (2.10)$$

Para as situações sem empates pode-se substituir a expressão central de (2.10) por:

$$\text{en} \quad \text{Re}(t) = \sum_{t_{e_j} \leq t} \frac{1}{n-j+1}. \quad (2.11)$$

Na tabela A7, no Anexo, são ilustrados os cálculos e determinadas as estimativas da curva de sobrevivência de Nelson, para os dados de transplante de rim. A curva é representada na figura 2.4., a seguir.

Figura 2.4 Estimativa da curva de sobrevivência (en) para os dados de transplante de rim, doador vivo esquema A.

A duração média estimada sob essa curva é de 1264,1 dias (somente 1,4% maior que a duração média estimada km) e a mediana estimada en é igual a 1503,3 (20,3% maior do que a estimativa bruta).

Ao contrário da curva (km), observa-se que, mesmo tendo no último ponto o registro de uma "morte"¹, a curva estimada (en) não chega, no final, a atingir o valor zero, embora ela fique mais baixa que a curva (km).

Da mesma forma como ocorre para as estimativas (km), também aqui as estimativas não levam em conta os tempos censurantes, mas o número de censuras em cada subintervalo I_m .

Uma consequência disto é a manutenção mais prolongada, na curva estimada, de patamares elevados nos subintervalos onde encontram-se as censuras, o que se reflete sensivelmente na estimativa da duração média ou mediana da sobrevida estudada.

2.4. O estimador exponencial por partes de Kitchin (ek)³⁴.

Este estimador não tem sido encontrado em aplicações tanto na área de saúde como na de confiabilidade. Ele apresenta basicamente a mesma idéia do estimador (en), no sentido da determinação da estimativa da curva de sobrevivência via estimativa da função risco acumulado. Entretanto, é matematicamente mais elaborado do que aquele, ajustando funções exponenciais a cada dois pontos consecutivos de "mortes^e", ao invés dos simples patamares de antes. Por isso, ele é identificado pela sigla "PEXE" que, em inglês, refere-se a "piece-wise exponential estimator".

Essa maior "sofisticação" matemática do estimador de Kitchin, quando comparado com o de Nelson ou 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.

O estimador de Kitchin não admite empates em cada tempo te_m , onde a taxa instantânea de "morte^e", definida em (1.3), é estimada por:

$$ek_{re_m} = \text{inverso da soma das áreas sob a curva } V[.] \text{ entre } te_{m-1} \text{ e } te_m^-, \quad (2.12)$$

A figura 2.5 ilustra este procedimento para uma situação hipotética com $E = \{1\}$ e $O = \{2\}$.

Figura 2.5 Estimação da taxa instantânea de risco r_1 , no m -ésimo subintervalo entre os tempos de duas "mortes" consecutivas.

A expressão geral do estimador de Kitchen é dada por:

$$\begin{aligned}
 {}^{ek}Se(t) = & \begin{cases} 1, & M(0,t) = 0 \text{ ou } t=0, \\ \exp \left\{ - \sum_{t_{e_{m-1}} < t} [(t_{e_{m-1}} - t_{e_{m-2}}) r_{e_{m-1}}] - (t - t_{e_{m-1}}) r_{e_m} \right\}, & \begin{cases} M(0,t) \geq 1, \\ t \in (t_{e_{m-1}}, t_{e_m}], \end{cases} \\ \exp \left\{ - \sum_{t_{e_m} \leq t} [(t_{e_m} - t_{e_{m-1}}) r_{e_m}] \right\}, & M(0,t) \geq 1 \text{ e } t > t_{e_m}, \end{cases}
 \end{aligned}
 \tag{2.13}$$

Note-se, na segunda linha da expressão (2.13), que, se t for um tempo maior do que $t_{e_{m-1}}$, a taxa de risco "acumulada no subintervalo I_m " (e que deve ser acrescida às taxas já acumuladas nos subintervalos anteriores, tão logo t atinja t_{e_m}), pode ser interpretada como o acréscimo da estimativa de r_{e_m} , em cada ponto deste subintervalo até o tempo t .

Por este estimador, as estimativas das probabilidades de sobrevivência ao risco Re decrescem em todos os pontos do subintervalo, anteriores ao tempo t_{e_m} , inclusive nos tempos de "censuras^o", aumentando a intensidade de decréscimo na "passagem" de um subintervalo para o próximo (conforme mostra a figura 2.6 a seguir).

Na última linha da expressão (2.13), observa-se que, caso

t não coincida com o último tempo de "morte"^e da amostra $t_{e_M} < t$, então a estimativa da probabilidade de sobrevivência ao risco R_e , nos tempos t, será a mesma que a calculada neste último ponto de "mortes"^e.

Em seu trabalho, Kitchin³⁴ provou as propriedades assintóticas de convergência fraca e forte deste estimador. As suas simulações indicaram em seu estimador "vício monotonamente crescente, substancialmente maior que o de Kaplan-Meier e menor erro quadrático médio, exceto nas caudas da suposta função teórica de sobrevivência".

Na tabela A8, no Anexo, é determinada a estimativa da correspondente curva de sobrevivência (representada na figura 2.6) para os dados de transplantes de rim. Note-se que, realmente, esse método exige muito mais trabalho computacional do que os anteriores.

Observe-se que, agora, a função de sobrevivência é representada por uma curva contínua composta, nos subintervalos $t_{1_{m-1}} < t \leq t_{1_m}$, por "pedaços" $s_m(t)$ de funções exponenciais definidas por:

$$s_m(t) = \exp\{-[R_1(t_{1_{m-1}}) + (t - t_{1_{m-1}})(r_{1_m})]\}. \quad (2.14)$$

Para a determinação da área sob essa curva deve-se recorrer à integração de cada um destes pedaços de funções exponenciais, entre os extremos dos correspondentes subintervalos e, depois, somar todos os resultados. Isto é:

$$\begin{aligned} \text{Área sob a curva kitchin} &= \sum_{m=1}^M \int_{t_{1_{m-1}}}^{t_{1_m}} s_m(t) dt. \\ &= \sum_{m=1}^M \exp\left\{-R_1(t_{1_{m-1}}) + r_{1_m} \cdot t_{1_{m-1}}\right\} \left[(1/r_{1_m}) \left\{\exp(-r_{1_m} \cdot t_{1_{m-1}}) - \exp(-r_{1_m} \cdot t_{1_m})\right\}\right] \end{aligned} \quad (2.15)$$

Conforme os cálculos na tabela 8A, a estimativa da duração média do tempo de sobrevida ao transplante de rim, determinada

pela área sob a trajetória da curva, é de 1187,9 dias (17,8% maior que a estimativa bruta e 4,9% menor que a (km).

Figura 2.6 Estimativa da curva de sobrevivência (ek) para os dados de transplante de rim, doador vivo, esquema A.

O cálculo da estimativa da mediana também torna-se aqui mais elaborado do que nas situações anteriores. Procura-se, na estimativa da curva de sobrevivência, o "pedaço exponencial" em cujo subintervalo esteja o tempo x' onde a curva, representada por esse "pedaço", assumo o valor 0,50. Isto é :

$$\begin{aligned}
 e_k \\
 S_1(x') &= 0,50 \Rightarrow \exp\{-[0,5878 + (x'-1424)(0,0011)]\} = 0,50 \Rightarrow \\
 \exp\{0,9786 - 0,0011 x'\} &= 0,50 \Rightarrow \exp\{0,9786\} \cdot \exp\{-0,0011 x'\} = 0,50 \Rightarrow
 \end{aligned}$$

$$-0,0011 \ x' = \text{Ln} [0,50 \exp(-0,9786)] \Rightarrow x' = 1519,8 \text{ dias.}$$

Portanto, a estimativa da mediana (ek) é 21,6 % maior que a estimativa bruta e aproximadamente igual à estimativa (km).

2.5 Comparação das estimativas da curva de sobrevivência segundo os estimadores F-N-P.

Finalizando este capítulo, comparam-se as estimativas da curva de sobrevivência, apresentando a figura 2.7, que resume os valores reunidos na tabela 2.1.

Figura 2.7 Comparação das estimativas (ea), (km), (en) e (ek) da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

Tabela 2.1 Estimativas (eb), (ea2), (km), (en) e (ek) da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

t	(eb)	(ea2)	(km)	(en)	(ek)
0	1,0000	1,0000	1,0000	1,0000	1,0000
40	0,9600	0,9600	0,9600	0,9608	0,9608
43	0,9200	0,9200	0,9200	0,9216	0,9216
89	0,8800	0,8800	0,8800	0,8824	0,8824
201	0,8400	0,8400	0,8400	0,8432	0,8432
215	0,8000	0,8000	0,8000	0,8039	0,8039
308	0,7600	0,7600	0,7600	0,7647	0,7647
396	0,7200	0,7200	0,7200	0,7255	0,7255
512	0,6800	0,6800	0,6800	0,6863	0,6863
542	0,6400	0,6400	0,6400	0,6471	0,6471
956*	0,6000	0,6400	0,6400	0,6471	0,6219
970*	0,5600	0,6400	0,6400	0,6471	0,6211
1227	0,5200	0,5973	0,5943	0,6025	0,6060
1272*	0,4800	0,5973	0,5943	0,6025	0,5941
1343*	0,4400	0,5973	0,5943	0,6025	0,5757
1362*	0,4000	0,5973	0,5943	0,6025	0,5709
1424	0,3600	0,5454	0,5349	0,5452	0,5555
1426*	0,3200	0,5454	0,5349	0,5452	0,5543
1434*	0,2800	0,5454	0,5349	0,5452	0,5494
1483*	0,2400	0,5454	0,5349	0,5452	0,5201
1518*	0,2000	0,5454	0,5349	0,5452	0,5001
1570	0,1600	0,4675	0,4279	0,4463	0,4719
1588*	0,1200	0,4675	0,4279	0,4463	0,4580
1609*	0,0800	0,4675	0,4279	0,4463	0,4436
1713*	0,0400	0,4675	0,4279	0,4463	0,3723
1973	0,0	0,2805	0,0	0,1642	0,2419
MÉDIA	1008,6	1215,6	1246,4	1264,1	1187,9
MÉDIA NA	1249,5	1485,1	1522,4	1503,3	1519,8

Obs: (eb) = estimativa bruta.

3. ALGUNS MÉTODOS NÃO PARAMÉTRICOS, SEGUNDO A INFERÊNCIA BAYESIANA, PARA ESTIMAR A CURVA DE SOBREVIVÊNCIA.

3.1. Conceitos básicos sobre a inferência bayesiana.

Uma outra ramificação teórica é a abordagem pela inferência bayesiana da estimação da curva de sobrevivência sob riscos competitivos^{64,70}. A linha básica de raciocínio da inferência bayesiana é inspirada em um artigo do Rev. Thomas Bayes, publicado, postumamente, em 1763⁹.

Uma das vantagens dos estimadores bayesianos, em relação aos seus correspondentes estimadores F-N-P, é a possibilidade de poder-se incorporar, em suas estimativas, algum conhecimento anterior pertinente aos dados sob estudo.

Suponha-se que Q seja uma quantidade sobre a qual se esteja incerto e que H seja o conjunto de conhecimentos (históricos), sobre os fenômenos relacionados a Q . Por exemplo: a quantidade desconhecida Q poderia ser a "verdadeira" função de sobrevivência $S_1(t)$ e o conjunto de conhecimentos históricos H poderiam ser as informações coletadas através de análises de seguimentos anteriores, semelhantes ao da situação atual, realizados ou não na própria UTR.

Diferentemente do que se convencionou na teoria F-N-P, considera-se, na inferência bayesiana, que essa quantidade Q não é fixa, mas aleatória. Deste modo, pode-se atribuir à Q uma distribuição de probabilidades a priori, condicionada aos conhecimentos H até então disponíveis, representada simbolicamente por $P[Q|H]$. Contrariamente ao que se faz na metodologia F-N-P, estas probabilidades são consideradas subjetivas, no sentido de dependerem da(s) fonte(s) anterior(es) de informações.

As informações extraídas dos dados amostrais D (que, no exemplo, são os tempos de duração em relação às datas de transplantes de rim e respectivas causas de "mortes"¹) são representadas na função de verossimilhança $v_D(Q) = P[D|Q]$ e incorporadas ao(s) conheci-

mento(s) anterior(es), por meio da chamada operação bayesiana (derivada do conhecido teorema de Bayes). Isto é, faz-se

$$P[Q|D \text{ e } H] = \frac{P[Q \text{ e } D \text{ e } H]}{P[D \text{ e } H]} = \frac{P[H] \times P[Q \text{ e } D|H]}{P[H] \times P[D|H]} = \frac{P[Q|H] \times P[D|Q \text{ e } H]}{P[D|H]}. \quad (3.1)$$

A partir do momento em que se conhece Q, a incerteza sobre D torna-se independente de qualquer conhecimento anterior H^{52} . Desse modo escreve-se $P[D|Q \text{ e } H] = P[D|Q]$ em (3.1) e obtem-se a expressão:

$$P[Q|D \text{ e } H] = \frac{P[Q|H] \times P[D|Q]}{P[D|H]}. \quad (3.2)$$

No primeiro membro de (3.2) tem-se a distribuição a posteriori, medindo a incerteza sobre Q após o conhecimento dos dados amostrais D e conhecimentos históricos H. Como o denominador do segundo membro dessa expressão é uma constante em relação a Q, ele pode ser "suprimido", para ser reescrita na forma de proporcionalidade:

$$P[Q|D \text{ e } H] \propto P[Q|H] \times P[D|Q]. \quad (3.3)$$

O símbolo $\propto$ é usado para indicar proporcionalidade. Observa-se, nesta expressão, que o primeiro fator do segundo membro é a distribuição a priori, representando a incerteza inicial sobre Q, isto é, antes da observação dos dados amostrais D.

Assim, a operação bayesiana consiste em se fazer:

$$\begin{array}{ccccc} P[Q|D \text{ e } H] & \propto & P[Q|H] & \times & v_D(Q). \\ \uparrow & & \uparrow & & \uparrow \\ \left[\begin{array}{c} \text{distribuição} \\ \text{a posteriori} \end{array} \right] & & \left[\begin{array}{c} \text{distribuição} \\ \text{a priori} \end{array} \right] & & \left[\begin{array}{c} \text{função de ve-} \\ \text{rossimilhança} \end{array} \right] \end{array} \quad (3.4)$$

Pode-se realizar este procedimento sequencialmente, renovando os conhecimentos sobre Q , por meio da atualização dos dados amostrais, obtidos com a realização de novos experimentos e utilizando a posteriori anterior como priori, no cálculo da próxima posteriori. A continuidade deste procedimento fica limitada a fatores tais como o tempo disponível para o estudo, os custos ou a qualidade dos conhecimentos disponíveis e/ou adquiridos.

Quando, no início de um tal processo, dispõe-se de nenhuma ou pouca informação a priori ou quando, simplesmente, deseja-se expressar objetividade na opinião sobre uma quantidade desconhecida Q , tenta-se selecionar uma distribuição a priori que melhor expresse essa "ignorância" sobre a quantidade em questão. Neste caso, diz-se que se está adotando uma "priori não informativa". Pode-se escolher, por exemplo, uma distribuição que não privilegie quaisquer das regiões do espaço amostral associado a Q .

Ao efetuar-se a operação bayesiana, pode-se ter situações analiticamente intratáveis. Para contornar isso, é conveniente utilizar-se uma distribuição a priori que, quando conectada aos dados experimentais via função de verossimilhança, gere uma distribuição a posteriori que pertença à mesma "família" da primeira. Por exemplo, quando os dados seguem um modelo multinomial (como em dados censurados de sobrevivência⁷⁰ semelhantes aos dos transplantes de rim), normalmente a distribuição de Dirichlet é utilizada como priori⁵².

Para decidir-se qual a melhor estimativa da quantidade Q de interesse, utiliza-se algum critério envolvendo o conceito de função perda. Conforme o próprio nome sugere, trata-se de estabelecer uma maneira de tentar-se quantificar as perdas nas quais se incorre ao optar por uma estimativa q , como avaliação dessa quantidade desconhecida que, na realidade, tenha Q como seu verdadeiro valor¹⁵. Uma função perda muito utilizada é a chamada função perda quadrática que, de uma forma muito simplificada, pode ser expressa por¹⁵:

$$D(Q, q) = (Q - q)^2. \quad (3.5)$$

Um critério razoável para determinar uma estimativa de Q é empregar a idéia de estimador de Bayes, isto é, considerar o valor q que minimize a perda esperada (também conhecida como função risco⁵² ou risco bayesiano⁷⁰). Este estimador é dado pela média condicional $E[Q|D \text{ e } H]$ da distribuição a posteriori $P[Q|D \text{ e } H]$.

As dificuldades para a implementação deste tipo de estimador são maiores que as apontadas para o estimador PEXE, seja porque existem diferenças de interpretação em relação ao que se pratica na inferência F-N-P ou porque quase sempre exigem maior elaboração nos cálculos. A necessidade do cumprimento de um maior número de etapas, algumas podendo incluir a resolução numérica de integrais, torna mais complexa a tarefa da construção de programas computacionais apropriados que forneçam de forma rápida, prática e confiável as estimativas das curvas de sobrevivência.

Serão apresentados, a seguir, dois estimadores bayesianos não paramétricos da função de sobrevivência sob dados censurados, estabelecidos assumindo-se como priori um processo de Dirichlet e o critério da função perda quadrática^{64,70}. O primeiro é o estimador de Suzarla e Van Ryzin que aqui é colocado, como recurso didático, para desempenhar o papel de etapa intermediária na transição dos estimadores F-N-P para o segundo estimador bayesiano, proposto por Salinas e Pereira em 1992.

3.2 O estimador de Suzarla e Van Ryzin (sv)⁶⁴.

Usando a função indicadora:

$$\delta_j = \begin{cases} 1, & \text{se } t_j = t_{e_m} \text{ (m-ésimo tempo de "morte")} \\ 0, & \text{se } t_j = t_{o_c} \text{ (c-ésimo tempo de censura)}, \end{cases} \quad (3.6)$$

a expressão básica do estimador de Kaplan e Meier, para a situação de não empates, pode ser escrita na forma (2.7) como⁴²

$$\begin{aligned}
\text{km} \\
\text{Se}(t) &= \prod_{t_j \leq t} \left(\frac{n-j}{n-j+1} \right)^{\delta_j} = \prod_{t_j \leq t} \left(\frac{n-j+1}{n-j} \right)^{-\delta_j} \\
&= \prod_{t_j \leq t} \left(\frac{n-j}{n-j+1} \right)^{-\delta_j} \cdot \left[\frac{1}{n} \left(\frac{n}{n-1} \cdot \frac{n-1}{n-2} \cdot \dots \cdot \frac{V(t)+1}{V(t)} \right) \frac{V(t)}{1} \right] \\
&= \frac{V(t)}{n} \prod_{t_j \leq t} \left(\frac{n-j+1}{n-j} \right)^{1-\delta_j}.
\end{aligned} \tag{3.7}$$

Para a situação de não empates, o estimador de Suzarla e Van Ryzin tem uma forma semelhante à (3.7), isto é:

$$\text{sv} \\
\text{Se}(t) = \frac{a(t)+V(t)}{a(0)+n} \prod_{t_j \leq t} \left(\frac{a(t_j)+(n-j+1)}{a(t_j)+(n-j)} \right)^{1-\delta_j}. \tag{3.8}$$

onde o parâmetro $a(t)$ é uma função contínua e monótona decrescente, no intervalo (t, ω) , que deve incorporar o conhecimento a priori no cálculo das estimativas. Por analogia, no caso de empates, o estimador é expresso por:

$$\text{sv} \\
\text{Se}(t) = \frac{a(t)+V(t)}{a(0)+n} \prod_{t_{o_c} \leq t} \left(\frac{a(t_{o_c})+V[t_{o_c}^-]}{a(t_{o_c})+V[t_{o_c}]} \right) \tag{3.9}$$

ou, considerando que o número de censuras no tempo t_{o_c} é dado por:

$$W[t_{o_c}] = V(t_{o_c}^-) - V(t_{o_c}),$$

$$\text{sv} \\
\text{Se}(t) = \frac{a(t)+V(t)}{a(0)+n} \prod_{t_{o_c} \leq t} \left(1 + \frac{W[t_{o_c}]}{a(t_{o_c})+V[t_{o_c}]} \right) \tag{3.10}$$

Se $a(t)$ convergir para zero, então o estimador de Suzarla e Van Ryzin torna-se uma versão modificada do correspondente estimador de Kaplan e Meier.

A estimativa a priori das probabilidades de sobrevivência, relativamente ao risco Re , pode ser expressa, segundo Miller(1981)⁴² ou Salinas e Pereira (1991)⁷⁰, pelo valor esperado a priori:

$$E[Se(t)] = \frac{a(t)}{a(0)} . \quad (3.11)$$

Uma vez definido o parâmetro $a(t)$ e dispondo das observações (t_i, C_i) , $i=1,2,\dots,n$, o cálculo das estimativas bayesianas de Suzarla e Van-Ryzin pode ser realizado sem nenhuma dificuldade.

3.3 Cálculo do estimador de Suzarla e Van Ryzin.

Como exemplo, estima-se a curva de sobrevivência associada aos dados aqui disponíveis sobre transplantes de rim. Para isso escolhe-se

$$a(t) = 1 - \frac{1}{U}t , \text{ com } U=2000 \text{ dias}$$

onde $U=2000$ foi um valor maior que $t_n = 1973$, fixado arbitrariamente, satisfazendo o seguinte critério:

$$P\{T_1 > t_n\} < 0,05. \quad (3.12)$$

Desse modo, tem-se $a(0)=1$ e $E[S_1(t)] = 1 - (0,0005)t$, isto é, está-se assumindo que a estimativa a priori da curva de sobrevivência ao risco 1 (morte do rim transplantado) seja uma função linear no intervalo $[0;2000]$, decrescendo de 1 (no ponto $t=0$) até atingir o valor zero (no ponto $t=2000$). Isso equivale a assumir que o tempo potencial T_1 de sobrevida do rim transplantado seja uniformemente distribuído no intervalo $[0;2000]$, isto é, quase nenhuma informação a

priori. Assim, por (3.9),

$$S_1(t) = \frac{(1-0,0005t)+V(t)}{1+25} \cdot \frac{(1-0,0005 t_2_c)+V[t_2_c]+1}{(1-0,0005 t_2_c)+V[T_2_c]}, \tag{3.13}$$

Os cálculos das estimativas são detalhados na tabela A9, no Anexo. A sua representação gráfica é apresentada na figura 3.1, junto com as estimativas das correspondentes curvas (km) e (ek), para facilitar a comparação visual. A comparação numérica pode ser feita utilizando-se a tabela 3.1, no final deste capítulo, onde foram reunidos os valores de todas essas estimativas.

Figura 3.1 Estimativas (km),(ek) e (sv), da curva de sobrevivência para os dados de transplante de rim, com doador vivo e tratamento A, admitindo-se $a(t)=1-0,0005 t$.

Como era de se esperar, existem poucas diferenças entre as estimativas (km) e (sv). Essa última apresentando leves inclinações nos "patamares", que vão se acentuando com o passar do tempo. No último ponto, que é um tempo de "morte"¹ a estimativa (sv) não chega a atingir o valor zero.

Nos subintervalos iniciais, sem as censuras, as estimativas (ek) são menores ou iguais às (sv). A partir dessas ocorrências as curvas se entrelaçam porque as estimativas (ek) nos pontos de mortes tornam-se maiores que as suas correspondentes (sv).

Conforme a tabela A9, a estimativa da média é 1227,4 dias [contra as estimativas (km) e (ek), de 1246,4 e 1187,9 dias, respectivamente]. A estimativa da mediana é, 1536,2 dias [contra as estimativas (km) e (ek), de 1522,4 e 1519,8 dias, respectivamente].

Segundo Miller⁴², Suzarla e Van Ryzin mostraram que, em muitos casos, o erro quadrático médio do seu estimador é menor do que o de Kaplan-Meier, mesmo quando se admite uma priori incorreta.

3.4 O estimador de Salinas e Pereira (sp)⁷⁰.

Supondo a não existência de pontos de descontinuidade (saltos) entre as funções de subsobrevivência $SE_T(.)$ e $SO_T(.)$ (onde, como antes, "E" e "O" estão se referindo aos conjuntos de "causas de interesse" e "outras causas", respectivamente) e usando a relação de Peterson⁵⁵, que expressa a função de sobrevivência $SE(.)$ em termos dessas funções de subsobrevivência, isto é:

$$SE(t) = \phi(SE_T(.), SO_T(.), t), \quad (3.14)$$

Salinas e Pereira chegaram ao seguinte estimador bayesiano:

$$SE(t) = S_T^{sp}(t) \cdot Y(t) \cdot \prod_{i=1}^{sp} \prod_{j=1}^{sp} I(t), \quad (3.15)$$

com

$$S_T^{sp}(t) = \frac{1}{\sum_{k=1}^K [ak(0)] + n} \left[\sum_{k=1}^K [ak(t)] + n S_T^{sp}(t) \right]; \quad (3.16)$$

$$Y(t) = \exp \left\{ \frac{-1}{\sum_{k=1}^K [ak(0)] + n} \left[\sum_{k \in O} \int_0^t \left(\frac{d[ak(u)]}{S_T^{sp}(u)} \right) \right] \right\}; \quad (3.17)$$

$$\prod_{t_0 \leq t}^{sp} = \prod_{t_0 \leq t} \left[1 + \frac{W[t_0]}{\sum_{k=1}^K ak(t_0) + V[t_0]} \right] \quad (3.18)$$

onde, como antes, os parâmetros $ak(t)$ são funções contínuas e monótonas decrescentes em (t, ω) , para $k=1, \dots, K$, relacionadas com a distribuição assumida a priori para representar uma previsão inicial sobre as probabilidades de sobrevivência.

Conforme já observado no trabalho de Salinas-Pereira, vê-se em (3.15) que, para $E=\{e\}$ e $O=\{o\}$, o produto

$$S_T(t) \cdot \prod(t) \quad (3.19)$$

torna-se similar ao estimador bayesiano de Susarla e Van Ryzin. Se todos os parâmetros $ak(t)$, $k=1, \dots, K$ convergirem para zero, a expressão (3.15) também se transforma em uma versão modificada do estimador de Kaplan e Meier. Salinas e Pereira também provaram as propriedades assintóticas⁷⁰ do seu estimador.

Nota-se que, ao contrário do estimador (sv), o estimador (sp) leva em conta em sua priori ambas as variáveis T_e e T_o .

3.5. Cálculo do estimador de Salinas e Pereira (sp).

A integral em (3.18) deve ser calculada mediante algum método de aproximação numérica. Por exemplo, Salinas-Pereira sugerem o "método Runge e Kutta de ordem 4".

Para isso, são considerados os distintos pontos ordenados t_j , $j = 1, 2, \dots, N$, de "mortes"^e ou de "censuras"^o mais o tempo $t_0 = 0$, isto é:

$$t_0, t_1, \dots, t_{j-1}, t_j, \dots, t_{N-1}, t_N$$

e, para cada $k \in O$, são definidas as funções

$$\Gamma_k(t) = \int_0^t \frac{1}{S_T^{SP}(u)} d[ak(u)] = \int_0^t \frac{1}{S_T^{SP}(u)} \frac{d}{du} [ak(u)] du = \int_0^t g_k(u) du. \quad (3.20)$$

Em cada subintervalo $I_j = (t_{j-1}, t_j]$, $j = 1, 2, \dots, N$, são estabelecidas as partições λ determinadas pelos $L-1$ pontos $z_1, \dots, z_{L-1}$ mais os extremos $z_0 = t_{j-1}$, $z_L = t_j$, conforme é ilustrado na figura (3.2):

Figura 3.2 Representação da partição λ em cada subintervalo I_j .

Nos pontos z_τ , $\tau = 0, 1, \dots, L$, a integral (3.20) é aproximada por:

$$\gamma_k(\tau) = \gamma_k(\tau-1) + \frac{\Delta}{6} [g_k(z_{\tau-1}) + 4g_k(z_{\tau-1} + \frac{\Delta}{2}) + g_k(z_{\tau-1} + \Delta)],$$

onde:

$$\gamma_k(0) = \Gamma_k(z_0), \quad (3.21)$$

$$\Gamma_k(0) = 0,$$

e

$$\Delta = z_\tau - z_{\tau-1} = \frac{t_j - t_{j-1}}{L}, \quad \tau = 1, 2, \dots, L.$$

Assim, uma aproximação numérica de (3.15) no subintervalo I_j é dada pelos pares:

$$\left\{ z_\tau; s_\tau^{sp}(z_\tau), y_\tau^{sp}(z_\tau), \Pi_\tau^{sp}(z_\tau) \right\}, \tau=0,1,2,\dots,L. \quad (3.22)$$

Para ilustração desses cálculos, utilizam-se os dados de transplantes de rim e, equivalentemente ao que foi feito no cálculo do estimador de Suzarla e Van Ryzin, define-se

$$a_1(t) = a_2(t) = 1 - \frac{t}{2000}, \text{ obtendo-se } a_1(0) = a_2(0) = 1. \quad (3.23)$$

Procedendo assim, conforme mostra Salinas⁷⁰, estar-se-á assumindo uma distribuição beta(1,1) [que é equivalente à distribuição uniforme(0,1)] como priori para a probabilidade $\rho = P[T_1 < T_2]$. Portanto, estar-se-á considerando uma priori não informativa para ρ .

Para simplificar os cálculos, considera-se os tempos variando dia a dia, de 0 até 1973, definindo as partições λ , em cada subintervalo I_j , como sendo constituídas pelos pontos correspondentes a cada dia dentro do subintervalo. Assim, faz-se $\Delta=1$ na expressão (3.21)

As estimativas das principais probabilidades de sobrevivência, bem como da média e mediana, segundo este estimador, são apresentadas na tabela 3.1, no final deste capítulo e comparadas com as correspondentes estimativas (km), (ek) e (sv). Na tabela A10, no Anexo, são apresentados os detalhes dos cálculos.

A representação gráfica (figura 3.3) é feita de forma similar à realizada para as estimativas (sv). Observa-se o mesmo comportamento apresentado na comparação das estimativas (sv), com sensível aumento nas diferenças agora. As estimativas da média e mediana são,

respectivamente, 1248.9 e 1518,3 dias (contra os 1246.4 dias e os 1522.4 dias da média e mediana (km), respectivamente.

Figura 3.3 Estimativas (km), (ek) e (sp) da curva de sobrevivência para os dados de transplante de rim, com doador vivo e tratamento A, admitindo-se uma priori não informativa uniforme para ρ e considerando $U1=U2=2000$ dias.

Tabela 3.1 Estimativas da curva de sobrevivência para os dados de transplante de rim, segundo os estimadores (km), (ek), (sv) [para $a(t) = 1 - 0,0005 t$] e (sp) [para $a_1(t) = 1 - 0,0005 t$ e $a_2(t) = 1 - 0,0005 t$].

t	km S1 (t)	ek S1 (t)	sv S1 (t)	sp S2 (t)
0	1,0000	1,0000	1,0000	1,0000
40	0,9600	0,9608	0,9608	0,9622
43	0,9200	0,9216	0,9222	0,9251
89	0,8800	0,8824	0,8829	0,8871
201	0,8400	0,8432	0,8423	0,8478
215	0,8000	0,8039	0,8036	0,8104
308	0,7600	0,7647	0,7633	0,7714
396	0,7200	0,7255	0,7231	0,7324
512	0,6800	0,6863	0,6825	0,6927
542	0,6400	0,6471	0,6434	0,6547
956*	0,6400	0,6219	0,6355	0,6469
970*	0,6400	0,6211	0,6352	0,6466
1227	0,5943	0,6060	0,5858	0,5976
1272*	0,5943	0,5941	0,5848	0,5967
1343*	0,5943	0,5757	0,5831	0,5950
1362*	0,5943	0,5709	0,5826	0,5945
1424	0,5349	0,5555	0,5244	0,5367
1426*	0,5349	0,5543	0,5244	0,5367
1434*	0,5349	0,5494	0,5241	0,5364
1483*	0,5349	0,5201	0,5223	0,5347
1518*	0,5349	0,5001	0,5209	0,5332
1570	0,4279	0,4719	0,4189	0,4330
1588*	0,4279	0,4580	0,4180	0,4321
1609*	0,4279	0,4424	0,4167	0,4308
1713*	0,4279	0,3723	0,4068	0,4213
1973	0,0	0,2419	0,0048	0,0099
Média $\bar{t}$	1246,4	1187,9	1227,4	1248,9
Mediana	1522,4	1519,8	1536,2	1518,3

DESENVOLVIMENTO TEÓRICO E RESULTADOS

4. EXPRESSANDO OS ESTIMADORES (F-N-P) EM TERMOS DE FUNÇÕES EMPÍRICAS.

4.1. Algumas aproximações discretas para o estabelecimento dos estimadores em termos de funções empíricas.

Considere-se, no seguimento, a atuação de somente dois riscos: o risco R_e de interesse e um outro risco R_o , onde incorporam-se todos os demais riscos não de interesse principal, mas que competem entre si e com R_e para a "retirada" dos indivíduos sob observação. Na notação adotada no capítulo 1, $E=\{e\}$ e $O=\{o\}=\{\text{todos os outros riscos (ou riscos censurantes)}\}$. Algumas vezes adota-se a simplificação $E \equiv e$ e $O \equiv o$. As taxas instantânea e acumulada, definidas em (1.3) e (1.2), respectivamente, são agora reescritas considerando o risco R_e , como:

$$r_e(t) = \frac{f_e(t)}{S_e(t)}, \quad 0 < t < \infty \quad (4.1)$$

$$e \quad R_e(t) = \int_0^t r_e(u) du, \quad 0 < t < \infty. \quad (4.2)$$

Estas taxas e o fato de que

$$S_e(t) = \exp(-R_e(t)), \quad 0 < t < \infty, \quad (4.3)$$

desempenham um papel importante na determinação da função de sobrevivência

$$S_e(t) = P[T_e > t], \quad 0 < t < \infty. \quad (4.4)$$

Visando a comparação analítica dos estimadores F-N-P estudados, as suas expressões serão agora reescritas em uma notação padronizada, em termos dos processos de contagem e funções empíricas,

definidos nos capítulos 1 e 2.

Estabelece-se

$$Se_T^n(te_m^-) - Se_T^n(te_m) = -[Se_T^n(te_m) - Se_T^n(te_m^-)] = \frac{1}{n} M[te_m], \quad (4.5)$$

e escreve-se

$$S_T^n(t_i) = \frac{1}{n} \sum_{j=1}^n I[T > t_i] = \frac{V[t_i]}{n}. \quad (4.6)$$

Essas expressões serão utilizadas no próximo item.

De Tsiatis-1975⁶⁷ tem-se que:

$$\frac{dSe_T(t)}{dt} = \frac{\delta S(te, to)}{\delta te} \Big|_{te=to=t}, \quad (4.7)$$

onde

$$S(te, to) = P[Te > te, To > to]. \quad (4.8)$$

A expressão (4.8) representa a probabilidade de que os tempos potenciais de sobrevida, submetidos aos riscos Re e Ro , excedam, respectivamente, os valores positivos te e to .

Admitindo, por hipótese, que:

[H1] - Os riscos Re e Ro sejam independentes,

então

$$S(te, to) = Se(te) \cdot So(to), \quad (4.9)$$

de onde vem, por (1.10) e (1.11),

$$Se(t) \cdot So(t) = S(t, t) = S_T(t) = Se_T(t) + So_T(t) \quad (4.10)$$

e, de (4.7) e (4.9),

$$\frac{dSe}{dt}_T(t) = \left[So(t_0) \cdot \frac{dSe(t_e)}{dt_e} \right]_{t_e=t_0=t} = So(t) \cdot \frac{dSe(t)}{dt} \quad (4.11)$$

Sendo $fe(t)$ a função densidade de probabilidade e $Fe(t)$ a função distribuição acumulada da variável aleatória Te , tem-se que

$$fe(t) = \frac{d}{dt} [Fe(t)] = - \frac{d}{dt} [1 - Fe(t)] = - \frac{d}{dt} [Se(t)] \quad (4.12)$$

e assim, de (4.1) a (4.3)

$$Se(t) = \exp \left[- \int_0^t \frac{fe(u)}{Se(u)} du \right] = \exp \left[\int_0^t \frac{\frac{d}{du} Se(u)}{Se(u)} du \right] \quad (4.13)$$

Multiplicando e dividindo a integranda da última igualdade de (4.13) pela função de sobrevivência aos riscos censurantes $So(u)$, aplicando em seguida (4.10) e (4.11) e da definição de diferencial, obtem-se, nessa sequência:

$$Se(t) = \exp \left[\int_0^t \frac{So(u) \frac{d}{du} Se(u)}{So(u) Se(u)} du \right] = \exp \left[\int_0^t \frac{\frac{dSe_T(u)}{du}}{S_T(u)} du \right] = \exp \left[\int_0^t \frac{d[Se_T(u)]}{S_T(u)} \right] \quad (4.14)$$

De (4.14), por analogia com (4.3) e (4.2), vê-se que

$$Se(t) = \exp - \left(\int_0^t \frac{-d[Se_T(u)]}{S_T(u)} \right) \quad (4.15)$$

de onde pode-se fazer a seguinte aproximação empírica de $Se(t)$:

$$Se_1^n(t) = \exp - \sum_{te_m \leq t} \frac{Se_T^n(te_m^-) - Se_T^n(te_m)}{S_T^n(te_m^-)}, \quad (4.16)$$

onde o índice 1 está sendo usado para diferenciar esta aproximação de outras a serem ainda estabelecidas neste capítulo. Assim,

$$Re_1^n(t) = \sum_{te_m \leq t} \frac{Se_T^n(te_m^-) - Se_T^n(te_m)}{S_T^n(te_m^-)} \quad (4.17)$$

onde, de (4.5) e (4.6), cada componente do somatório pode ser expresso por:

$$re_1^n(I_m) = \frac{Se_T^n(te_m^-) - Se_T^n(te_m)}{S_T^n(te_m^-)} = \frac{M[te_m]}{V[te_m^-]}, \quad (4.18)$$

que é a aproximação empírica "do acúmulo da taxa $re(.)$ no subintervalo I_m , que vai de te_{m-1} até te_m ". No próximo capítulo esse assunto será tratado com mais detalhe, apresentando uma interpretação geométrica para essas componentes.

Retomando, por um momento, o raciocínio desenvolvido até (4.15) e admitindo, além da hipótese [H1], que:

[H2]- As funções de sobrevivência $Se(.)$ e $So(.)$ não têm pontos de salto em comum,

então $Se_T(.)$ e $So_T(.)$ também não terão pontos de salto em comum (conforme Peterson-1977)⁵⁵.

Nestas condições, como te_m é considerado ponto de salto somente de $Se(.)$, da primeira igualdade de (4.18), pode-se escrever

$$re_1^n(I_m) = \frac{[Se_T^n(te_m^-) + So_T^n(te_m^-)] - [Se_T^n(te_m) + So_T^n(te_m)]}{S_T^n(te_m^-)}$$

e, por (4.10), expressar $re_1^n(I_m)$ somente em relação à função empírica $S_T^n(.)$

$$re_1^n(I_m) = \frac{S_T^n(te_m^-) - S_T^n(te_m)}{S_T^n(te_m^-)} \tag{4.19}$$

4.2. Expressando o estimador de Nelson em termos de funções empíricas.

Procura-se um estimador cuja expressão tenha basicamente a forma (4.16). Os estimadores de Nelson e de Kitchen são, dentre os a apresentados, os que mais diretamente aproximam-se desse modelo. A maior simplicidade na expressão matemática do primeiro, naturalmente leva a iniciar a investigação de qual a sua estrutura em relação às funções empíricas.

De (4.16) e (4.18) tem-se:

$$Se_1^n(t) = \exp \sum_{te \leq t} \frac{M[te_m]}{V[te_m^-]} = Se(t) \tag{4.20}$$

Assim, $Se_1^n(t)$ é justamente o estimador de Nelson, para a curva de sobrevivência. Note-se que a sua taxa acumulada é estimada

em termos das funções empíricas $S_T^n(\cdot)$ e $p^n = \frac{M[\cdot]}{n}$, isto é:

$$Se(t) = \exp - \sum_{te_m \leq t} \frac{M[te_m]}{n S_T^n[te_m^-]}. \quad (4.21)$$

4.3. Expressando o estimador de Kaplan e Meier em termos de funções empíricas.

De (4.16) e primeira igualdade de (4.18)

$$Se_1^n(t) = \exp - \sum_{te_m \leq t} re_1^n(I_m). \quad (4.22)$$

Agora, em (4.13) faz-se a seguinte substituição:

$$\frac{1}{Se(u)} \frac{d}{du} Se(u) du = \frac{d}{du} \ln[Se(u)] du = d\{\ln[Se(u)]\} \quad (4.23)$$

e escreve-se:

$$Se(t) = \exp \int_0^t d\{\ln[Se(u)]\}, \quad (4.24)$$

que é aproximada empiricamente por:

$$Se_2^n(t) = \exp \sum_{te_m \leq t} [\ln Se^n(te_m) - \ln Se^n(te_m^-)] = \exp \sum_{te_m \leq t} \ln \left[\frac{Se^n(te_m)}{Se^n(te_m^-)} \right]. \quad (4.25)$$

Mas,

$$\begin{aligned}
\frac{Se^n(te_m)}{Se^n(te_m^-)} &= \frac{Se^n(te_m).So^n(te_m^-)}{Se^n(te_m^-).So^n(te_m^-)} = \frac{Se^n(te_m).So^n(te_m^-)}{Se^n(te_m^-).So^n(te_m^-)} = \frac{S_T^n(te_m)}{S_T^n(te_m^-)}, \\
&\quad \uparrow \quad \quad \quad \uparrow \\
&\quad \text{[por H2]} \quad \quad \quad \text{[por (4.10)]} \\
&= \frac{S_T^n(te_m^-) - (S_T^n(te_m^-) - S_T^n(te_m))}{S_T^n(te_m^-)} = 1 - \frac{S_T^n(te_m^-) - S_T^n(te_m)}{S_T^n(te_m^-)}
\end{aligned}
\tag{4.26}$$

De (4.26) e (4.19):

$$\frac{Se^n(te_m)}{Se^n(te_m^-)} = 1 - re_1^n(I_m)
\tag{4.27}$$

e portanto a última igualdade de (4.25) pode ser escrita na forma:

$$Se_2^n(t) = \exp \sum_{te_m \leq t} \ln[1 - re_1^n(I_m)].
\tag{4.28}$$

Substituindo a segunda igualdade da expressão (4.18) no resultado anterior, tem-se

$$Se_2^n(t) = \exp \sum_{te_m \leq t} \ln \left[1 - \frac{M[te_m]}{V[te_m^-]} \right] = \exp \ln \prod_{te_m \leq t} \left[1 - \frac{M[te_m]}{V[te_m^-]} \right],
\tag{4.29}$$

e assim

$$Se_2^n(t) = \prod_{te_m \leq t} \left[1 - \frac{M[te_m]}{V[te_m^-]} \right]^{km} = Se(t),
\tag{4.30}$$

que é o estimador de Kaplan-Meier. Na forma análoga a (4.21) tem-se

$${}^{km}Se(t) = \exp - \sum_{te_m \leq t} \ln \left[\frac{S_T^n(te_m^-)}{S_T^n(te_m)} \right]. \quad (4.31)$$

Note-se que, por analogia com (4.22) e de (4.28),

$${}^{km}re(I_m) = -\ln[1 - re_1^n(I_m)]. \quad (4.32)$$

4.4. Expressando o estimador atuarial em termos de funções empíricas.

De (2.3) e (2.4), o estimador atuarial pode ser escrito na forma:

$${}^{ea}Se(t) = \prod_{te_m \leq t} \left[1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} \right]. \quad (4.33)$$

Por analogia com (4.29) e fazendo a seguinte substituição:

$$1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} = \frac{V[te_m] + \frac{W[I_m]}{2}}{V[te_m^-] + \frac{W[I_m]}{2}} \quad (4.34)$$

escreve-se

$${}^{ea}Se(t) = \exp - \sum_{te_m \leq t} \ln \left[\frac{nS_T^n(te_m^-) + \frac{W[I_m]}{2}}{nS_T^n(te_m) + \frac{W[I_m]}{2}} \right], \quad (4.35)$$

Note-se que, de (4.33):

$${}^{ea}Se(t) = \exp - \sum_{te_m \leq t} -\ln \left[1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} \right] \quad (4.36)$$

e por analogia com (4.22),

$$Ea_{re}(I_m) = -\ln \left[1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} \right] \quad (4.37)$$

4.5. Expressando o estimador de Kitchin em termos de funções empíricas.

Com base em argumentos geométricos, será mostrado, no próximo capítulo, que a expressão do estimador de Kitchin em termos de funções empíricas, nos pontos te_s ($s \geq m$), de "mortes^e", (isto é, da parte discreta do estimador, pois ele é contínuo), é a seguinte:

$$Ea_{Se}(te_s) = \exp - \sum_{te_m \leq te_s} \frac{M[te_m]}{nS_T^n(te_m^-) + \sum_{c=1}^{W[I_m]} f_c(I_m)}, \quad (4.38)$$

onde:

- 1- c é o índice relacionado com os pontos de censura,
- 2- $f_c(I_m)$ representa a fração do comprimento de I_m que corresponde à distância da c -ésima censura deste subintervalo em relação ao ponto inicial te_{m-1} .

A rigor, $M[te_m]$ é sempre igual a 1, para o estimador de Kitchin. Esta notação está sendo usada para facilitar a comparação analítica desta com as expressões equivalentes dos demais estimadores.

5. COMPARAÇÃO ENTRE OS ESTIMADORES DAS TAXAS $re(I_m)$.

5.1. Interpretação geométrica das estimativas de $re(t)$.

De maneira análoga à abordagem de Kitchin, interpreta-se a gora a taxa instantânea $re_m(te_m)$, no estimador de Nelson, como sendo o inverso da área sob a trajetória do processo de contagem $V[.]$, entre os extremos te_{m-1} e te_m de um subintervalo I_m . As diferenças entre os valores dessas taxas, quando se usa um ou outro desses estimadores, surgem basicamente a partir da ocorrência das censuras, conforme o critério de escolha da trajetória de $V[.]$.

Na figura 5.1 ilustra-se isso, para dois subintervalos I_m , fixados em uma situação hipotética.

Figura 5.1 Representação das áreas, consideradas na interpretação da estimação das taxas $re(I_m)$, segundo os estimadores (ek) e (en) , em uma situação hipotética.

Na figura 5.1, apresenta-se as trajetórias $V[.]$ (indicadas pelas linhas contínuas mais espessas), consideradas na interpretação geométrica da estimativa das taxas $re(I_m)$, segundo o estimador (ek) [vejam-se figuras (i) e (ii)] e segundo o estimador (en) [vejam-se figuras (iii) e (iv)]. As áreas sob elas são representadas recorrendo-se ao seguinte expediente:

$$A_{ek} = \text{Área formada por todos os retângulos assinalados pela letra maiúscula A [veja-se efeito das censuras, na figura (ii)]} \quad (5.1)$$

$$= V[te_{m-1}][to_1 - te_{m-1}] + V[to_1][to_2 - to_1] + V[te_m^-][te_m - to_2],$$

$$A_{en} = \text{Área formada por todos os retângulos assinalados pela letra minúscula a [veja-se efeito das censuras, na figura (iv)]} \quad (5.2)$$

$$= V[te_m^-][te_m - te_{m-1}],$$

$$\text{e } A_b = \text{Área formada por todos os retângulos assinalados pela letra minúscula b [veja-se efeito das "mortes" nas figuras (ii) e (iv)]} \quad (5.3)$$

$$= M[te_m][te_m - te_{m-1}].$$

Na aproximação de Kitchin, a contribuição de cada censura para a área (e conseqüentemente para a taxa) é computada até o ponto de sua ocorrência no subintervalo, fazendo $V[.]$ descrever uma trajetória com a formação de "degraus", tanto nos pontos de censuras, como nos pontos de "mortes". Na aproximação de Nelson, esses elementos de áreas correspondentes às permanências dos indivíduos sob seguimento até o seu ponto de censura, são suprimidos logo no início do subintervalo. Assim, a trajetória de $V[.]$ torna-se constante no subintervalo, só apresentando (grandes) saltos nos seus extremos. Em conseqüência, as áreas A_{en} são iguais ou menores que as correspondentes áreas A_{ek} .

5.2. Estimativas da taxa $re(I_m)$.

Suponha-se inicialmente que $M[te_m] = 1$ e interprete-se a estimativa (en) da taxa instantânea $re(.)$, no ponto te_m , como sendo o inverso da área sob a trajetória constante $V[te_m^-]$, entre os extremos de I_m . Então, multiplicando e dividindo o termo dentro do somatório de (4.20) pelo seguinte comprimento de subintervalo:

$$\Delta te_m = [te_m - te_{m-1}], \quad (5.4)$$

tem-se

$$Se(t) = \exp - \sum_{te_m \leq t} \Delta te_m re^{en}(te_m). \quad (5.5)$$

Assim, define-se o termo dentro do somatório, em (5.5), como a estimativa de Nelson da taxa acumulada no subintervalo $re(I_m)$, daqui em diante simplesmente denominada "taxa no subintervalo".

Pode-se estender esse raciocínio para as situações envolvendo empates em pontos de "mortes"^e.

De uma maneira mais geral, define-se

$$re(I_m)_{xx} = M[te_m] \Delta te_m re^{xx}(te_m) = \frac{A_b}{A_{xx}}, \quad (5.6)$$

onde xx é a sigla do estimador em referência: en para Nelson, ek para o de Kitchin (embora, pela continuidade deste último, só possa haver $M[te_m] = 1$). Adota-se a notação padrão $M[te_m]$, mesmo para o estimador de Kitchin, visando comparações analíticas entre as expressões dos estimadores das taxas acumuladas definidas por (5.6). Esse tipo de associação de taxas com áreas, além de simplificar a representação e permitir generalizações, também facilita comparações analíticas, conforme ver-se-á a seguir.

5.3. Comparação das estimativas das taxas $re(I_m)$, considerando os estimadores F-N-P apresentados e interpretação geométrica.

Na procura de um estimador F-N-P, alternativo aos apresentados, para a curva de sobrevivência, optou-se pela definição de sua taxa, com base no conhecimento do comportamento das taxas dos estimadores comentados no capítulo 4.

Assim, fixando um subintervalo genérico I_m e utilizando interpretação por áreas, compara-se, a seguir, as estimativas das taxas $re(I_m)$ daqueles estimadores. Isto equivale a comparar a "parte discreta" das correspondentes funções de sobrevivência neste subintervalo.

5.3.1. Comparação entre as taxas (en) e (ek) .

Para os estimadores (en) e (ek) essa comparação é imediata, pois do fato de

$$A_{ek} \geq A_{en},$$

tem-se

$$re(I_m)^{ek} = \frac{A_b}{A_{ek}} \leq \frac{A_b}{A_{en}} = re(I_m)^{en}. \quad (5.7)$$

A igualdade em (5.7) é válida quando não ocorrem pontos de censura no subintervalo I_m . Vê-se, então, que para um mesmo subintervalo I_m , a estimativa da taxa $re(I_m)$, produzida pelo estimador (en) , é maior ou igual que a estimativa dada pelo estimador (ek) .

Note-se que, na "força de mortalidade ao risco Re ", no ponto te_m , avaliada pelo estimador de Kitchin, dá-se um maior "desconto", por conta dos efeitos dos riscos competitivos censurantes Ro , do que na correspondente "força de mortalidade ao risco Re ", pelo estimador de Nelson.

Uma das vantagens da estimativa (en) é a sua maior simplicidade de cálculo e facilidade de interpretação, em relação às demais estimativas.

5.3.2. Interpretação da taxa (km) em termos de áreas.

Na extensão dessa interpretação geométrica das taxas, uma pergunta que surge é: como interpretar a estimativa representada por (4.32), em termos de áreas sob a trajetória do processo de contagem $V[.]$?

Para verificar isso, desenvolve-se $re(I_m^{km})$, a partir da definição apresentada em (4.32) e desenha-se a figura 5.2 a seguir, isto é:

$$\begin{aligned}
 km \\
 re(I_m) &= -\ln \left[1 - \frac{M[te_m^-]}{V[te_m^-]} \right] = -\ln \left[1 - \frac{M[te_m^-] \Delta te_m}{V[te_m^-] \Delta te_m} \right] \\
 &= -\ln \left[\frac{V[te_m^-] \Delta te_m - M[te_m^-] \Delta te_m}{V[te_m^-] \Delta te_m} \right] = -\ln \left[\frac{V[te_m^-] \Delta te_m}{V[te_m^-] \Delta te_m} \right].
 \end{aligned}
 \tag{5.8}$$

Ou, então:

$$km \\
 re(I_m) = \ln V[te_m^-] - \ln V[te_m].
 \tag{5.9}$$

Figura 5.2 Representação das áreas, consideradas na interpretação da estimação das taxas $re(I_m)$, segundo o estimador (km).

Adicionalmente às áreas já definidas anteriormente, define-se a seguinte:

$$A_B = \text{Área formada pelos retângulos assinalados pela letra maiúscula B (conforme a figura 5.2)} \quad (5.10)$$

$$= V[te_m] \Delta te_m.$$

Da terceira igualdade de (5.8):

$$re(I_m^{km}) = -\ln \left[\frac{A_{en} - A_B}{A_{en}} \right] = -\ln \left[\frac{A_B}{A_{en}} \right] = \ln(A_{en}) - \ln(A_B). \quad (5.11)$$

Portanto, a estimativa da taxa $re(I_m)$, segundo o estimador de Kaplan e Meier, pode ser interpretada como uma diferença de logarítmos de áreas do mesmo tipo já definido anteriormente.

5.3.3. Interpretação da taxa (ea) em termos de áreas.

Para interpretar a taxa $re(I_m)$ no estimador atuarial, considera-se as seguintes áreas:

$$A_{ea} = V[te_{m-1}] \Delta te_m, \quad (5.12)$$

$$\bar{A} = \frac{A_{ea} + A_{en}}{2}. \quad (5.13)$$

Note-se que:

$$\bar{A} = \frac{(V[te_{m-1}] + V[te_m^-]) \Delta te_m}{2} = \frac{(V[te_{m-1}] + V[te_{m-1}] - W[I_m]) \Delta te_m}{2}$$

e, portanto:

$$\bar{A} = \left\{ V[te_{m-1}] - W[I_m]/2 \right\} \Delta te_m. \quad (5.14)$$

De (4.37) e (2.4) tem-se:

$$\begin{aligned} {}^{ea}re(I_m) &= -\ln \left[1 - \frac{M[te_m] \Delta te_m}{(V[te_{m-1}] - \frac{W[I_m]}{2}) \Delta te_m} \right] \\ &= -\ln \left[1 - \frac{A_b}{\bar{A}} \right] = -\ln \left[\frac{\bar{A} - A_b}{\bar{A}} \right] \end{aligned} \quad (5.15)$$

5.3.4 Comparação entre taxas (km) e (ea).

De (4.34) e (4.37)

$${}^{ea}re(I_m) = \ln \left\{ V[te_m] + \frac{W[I_m]}{2} \right\} - \ln \left\{ V[te_m] + \frac{W[I_m]}{2} \right\}, \quad (5.16)$$

que pode ser comparada com a expressão (5.9).

Como:

$$\ln x - \ln y \geq \ln(x+c) - \ln(y+c), \text{ para } c > 0, \quad (5.17)$$

por (5.9), (5.16) e (5.17) tem-se

$${}^{ea}re(I_m) \leq {}^{km}re(I_m), \quad (5.18)$$

onde a desigualdade tende a acentuar-se com o aumento na quantidade $W[I_m]$ de censuras no subintervalo I_m .

5.3.5. Comparação entre as taxas (ek), (en) e (km).

Da seguinte desigualdade (Thomas-1972)⁶⁶,

$$0 < -\ln[1 - (x+1)^{-1}] - (x+1)^{-1} < [x(x+1)]^{-1}, \quad \text{para } x > 0, \quad (5.19)$$

escreve-se:

$$\frac{1}{x+1} < -\ln\left(1 - \frac{1}{x+1}\right) < \frac{1}{x}. \quad (5.20)$$

Para $V[te_m] > 0$ e $M[te_m] > 0$, pode-se fazer $x = \frac{V[te_m]}{M[te_m]}$,

obtendo-se

$$\frac{M[te_m]}{V[te_m]} < -\ln\left(1 - \frac{M[te_m]}{V[te_m]}\right) < \frac{M[te_m]}{V[te_m]}. \quad (5.21)$$

Da definição (5.6) aplicada ao estimador de Nelson, da primeira igualdade de (5.8) e da primeira desigualdade de (5.21) tem-se

$$\frac{en}{re(I_m)} < \frac{km}{re(I_m)} \quad (5.22)$$

e de (5.7)

$$\frac{ek}{re(I_m)} \leq \frac{en}{re(I_m)} < \frac{km}{re(I_m)}. \quad (5.23)$$

5.3.6. Comparação entre as taxas (en), (km).

Se, para $M[te_m] > 0$ e $W[I_m] > 0$, definir-se em (5.20),

$$x = \frac{V[te_m] + \frac{W[I_m]}{2}}{M[te_m]}$$

$$\frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} < -\ln \left(1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} \right) < \frac{M[te_m]}{V[te_m] + \frac{W[I_m]}{2}}. \quad (5.24)$$

Por (2.4), (5.3) e (5.14) o fator extremo esquerdo da des-
sa desigualdade pode ser escrito como:

$$\frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} = \frac{M[te_m]}{V[te_{m-1}] - \frac{W[I_m]}{2}} = \frac{A_b}{\bar{A}}, \quad (5.25)$$

que é o fator atuarial que estima as probabilidades de "mortes^e" no subintervalo I_m , quando considera-se a forma (2.3). O seu papel na composição da taxa atuarial é equivalente ao que o estimador de Nelson da taxa $re(I_m)$ desempenha na composição do correspondente estimador de Kaplan e Meier [veja-se (4.37) e (4.32)]. Além disso, a razão de áreas na última igualdade sugere uma interpretação geométrica análoga às das taxas de Nelson e de Kitchin.

Assim, por (5.25), pode-se definir uma "taxa padrão":

$$ep_{re(I_m)} = \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} = \frac{A_b}{\bar{A}}$$

De (5.13) e (5.2) tem-se que $\bar{A} \geq A_{en}$. Portanto,

$$ep_{re(I_m)} = \frac{A_b}{\bar{A}} \leq \frac{A_b}{A_{en}} = en_{re(I_m)}.$$

e, por (5.23),

$$ep_{re(I_m)} \leq en_{re(I_m)} < km_{re(I_m)}. \quad (5.26)$$

5.3.7. Comparação entre as taxas (ep), (ea) e (km).

Da primeira desigualdade de (5.24),

$$\begin{matrix} ep & ea \\ re(I_m) & < re(I_m); \end{matrix} \quad (5.27)$$

com (5.18), tem-se

$$\begin{matrix} ep & ea & km \\ re(I_m) & < re(I_m) \leq re(I_m). \end{matrix} \quad (5.28)$$

Até este ponto, com (5.23), (5.26) e (5.28), compararam-se as taxas acumuladas em I_m , envolvendo os estimadores (ek), (en), (km) e (ep), obtendo ordenações, aqui simplificadas por:

$$\begin{array}{rclclcl} ek & \leq & en & < & km \\ ep & \leq & en & < & km \\ ep & < & ea & \leq & km. \end{array} \quad (5.29)$$

O exame de (5.29) leva à formulação das seguintes perguntas. Em um mesmo subintervalo I_m :

1- qual a posição da taxa (ek), em relação às correspondentes taxas (ea) e (ep) ?

2- qual a posição da taxa (en), em relação à taxa (ea) ?

5.3.8. Comparação entre as taxas (ep) e (ek).

Para facilitar a compreensão do raciocínio empregado na busca das respostas às perguntas formuladas acima, ilustram-se quatro situações particulares, nas figuras 5.3, representando possíveis comportamentos da trajetória do processo $V[.]$, durante o acompanhamento em um seguimento com dados censurados.

Figura 5.3 Quatro possíveis situações sobre o comportamento do processo $V[.]$, em um subintervalo I_m .

Observa-se, pela situação (i), que se não ocorrerem censuras em I_m , então $A_{ek} \equiv \bar{A}$. Isso também acontece quando todas as censuras ocorrerem na metade de I_m (no ponto M), ou ainda, conforme é ilustrado na situação (ii), quando os tempos de censura forem equidistantes em relação à metade de I_m . Na situação (iii) tem-se as censuras ocorrendo em tempos posicionados mais para o início do subintervalo e em consequência $A_{ek} < \bar{A}$. Na situação (iv) tem-se o contrário e, portanto $A_{ek} > \bar{A}$.

De um modo geral:

1 - quando $\bar{A} \leq A_{ek}$, ter-se-á

$$\text{ep} \text{ re}(I_m) = \frac{A_b}{\bar{A}} \geq \frac{A_b}{A_{ek}} = \text{ek} \text{ re}(I_m), \quad (5.30)$$

2 - quando $\bar{A} \geq A_{ek}$, ter-se-á

$$re(I_m^{ep}) = \frac{A_b}{\bar{A}} \leq \frac{A_b}{A_{ek}} = re(I_m^{ek}). \quad (5.31)$$

Vê-se que em um mesmo segmento podem ocorrer alguns subintervalos em que as taxas (ep) e (ek) satisfazem a ordenação (5.31), enquanto em outros ocorre o contrário. Portanto, aqui podem ocorrer situações em que a ordenação não siga estritamente o mesmo sentido.

5.3.9. Definição de elementos de área auxiliares para comparar as taxas (ek) e (en) , em relação à taxa (ea) .

No próximo item, serão investigadas as posições das taxas (ek) e (en) , em relação à taxa (ea) . Para isso serão usados, respectivamente, dois elementos de área definidos, para $\bar{A} \geq A_{ek}$ por:

$$\Delta A = \bar{A} - A_{ek}, \quad (5.32)$$

$$e \quad \Delta B = \bar{A} - A_{en}. \quad (5.33)$$

cujas representações gráficas, para uma situação particular, encontram-se na figura 5.4.

Figura 5.4 Representações gráficas, para uma situação particular, dos elementos de área ΔA e ΔB .

Note-se, na figura 5.4.(i), que o "deslocamento" do elemento de área hachurada facilita a visualização da diferença $\bar{A} - A_{ek}$, que produz ΔA . Assim, a área $\bar{A} - \Delta A$ corresponde à área A_{ek} .

5.3.10. Comparação entre as taxas (ek) e (ea).

Para verificar a posição da taxa (ek), em relação à taxa (ea), usa-se um raciocínio semelhante ao anterior.

1 - quando $\bar{A} \leq A_{ek}$,

$$re(I_m^{ea}) = -\ln\left(1 - \frac{A_b}{\bar{A}}\right) > \frac{A_b}{\bar{A}} \geq \frac{A_b}{A_{ek}} = re(I_m^{ek}),$$

e, portanto, nestas condições:

$$re(I_m^{ea}) \geq re(I_m^{ek}). \quad (5.34)$$

2 - quando $\bar{A} > A_{ek}$, deve-se elaborar um pouco mais, utilizando o elemento de área definido em (5.32).

Para $\frac{1}{x} = \frac{\Delta A}{\bar{A} - \Delta A}$ em (5.20), $A_b \leq \Delta A$ e (5.32) escreve-se

$$\frac{A_b}{\bar{A}} < -\ln\left(1 - \frac{A_b}{\bar{A}}\right) \leq -\ln\left(1 - \frac{\Delta A}{\bar{A}}\right) < \frac{\Delta A}{\bar{A} - \Delta A} = \frac{\Delta A}{A_{ek}} \quad (5.35)$$

e na situação particular em que $A_b = \Delta A$ tem-se:

$$re(I_m^{ea}) = -\ln\left(1 - \frac{A_b}{\bar{A}}\right) < \frac{\Delta A}{A_{ek}} = \frac{A_b}{A_{ek}} = re(I_m^{ek}) \quad (5.36)$$

Vê-se, então, que em pelo menos uma situação particular, a desigualdade entre as taxas (ea) e (ek) tem o sentido contrário ao da desigualdade apresentada em (5.34).

5.3.11. Comparação entre as taxas (en) e (ea).

Para verificar-se a posição da taxa (en), em relação à taxa (ea), procede-se de maneira análoga à da situação anterior.

Utilizando o elemento de área definido em (5.33), faz-se:

$$\frac{1}{x} = \frac{\Delta B}{\bar{A} - \Delta B}$$

em (5.20) e para $A_b \leq \Delta B$ escreve-se:

$$\frac{A_b}{\bar{A}} < -\ln\left(1 - \frac{A_b}{\bar{A}}\right) \leq -\ln\left(1 - \frac{\Delta B}{\bar{A}}\right) < \frac{\Delta B}{\bar{A} - \Delta B} = \frac{\Delta B}{A_{en}}. \quad (5.37)$$

1 - Na situação particular em que $A_b = \Delta B$, tem-se:

$${}^{ea}_{re}(I_m) = -\ln\left(1 - \frac{A_b}{\bar{A}}\right) < \frac{\Delta B}{A_{en}} = \frac{A_b}{A_{en}} = {}^{en}_{re}(I_m). \quad (5.38)$$

2 - Quando não ocorrem censuras em I_m , tem-se $\bar{A} = A_{en}$, isto é

$${}^{ea}_{re}(I_m) = -\ln\left(1 - \frac{A_b}{\bar{A}}\right) > \frac{A_b}{\bar{A}} = \frac{A_b}{A_{en}} = {}^{en}_{re}(I_m). \quad (5.39)$$

Então, vê-se, por (5.38) e (5.39), que a ordem de desigualdade das taxas (ea) e (en) pode mudar de sentido, dependendo da situação em cada subintervalo I_m .

5.4. Expressão da parte discreta do estimador de Kitchin.

Conforme já adiantado no capítulo anterior, o estimador de Kitchin pode ser expresso somente nos pontos de "mortes^e", o que pode ser chamado de sua "parte discreta". Explicar-se-á, agora, a construção dessa expressão, com base na interpretação geométrica da sua taxa $re(I_m)$ e serão utilizadas figuras para representar os elementos de área envolvidos, considerando algumas situações específicas.

Em cada caso, supõe-se um número fixo $M[te_m]$ de "mortes^e" em te_m e um número $W[I_m]$ de "censuras^o" (variando de zero a dois) no subintervalo I_m . Inclusive, para facilitar a comparação analítica da expressão geral de sua taxa $re(I_m)$ com as dos demais estimadores F-N-P, serão considerados empates nos pontos (o que teoricamente não é permitido pois trata-se de estimador contínuo). As censuras são alocadas em tempos situados nas seguintes posições: no ponto correspondente à metade de I_m ; nos pontos correspondentes ao primeiro e terceiro quartos de I_m e nos pontos correspondentes ao primeiro, terceiro, quinto e sétimo oitavos de I_m . Nas figuras 5.5, de (i) até (xii), estão representadas as distribuições das mortes e censuras, os elementos geométricos e as estimativas das taxas.

Figura 5.5 Distribuição das "censuras^o" e "mortes^e" no subintervalo I_m .

nenhuma censura

Figura 5.5-i

$$\frac{ek}{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m}$$

uma censura

Figura 5.5-ii

$$\frac{ek}{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \frac{1}{2} \cdot \Delta te_m \cdot 1}$$

uma censura

Figura 5.5-iii

$$\frac{ek}{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \frac{1}{4} \cdot \Delta te_m \cdot 1}$$

uma censura

Figura 5.5-iv

$$\frac{ek}{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \frac{3}{4} \cdot \Delta te_m \cdot 1}$$

Figura 5.5-v

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + (\frac{1}{2} + \frac{1}{2}) \Delta te_m}$$

Figura 5.5-vi

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + (\frac{1}{4} + \frac{3}{4}) \Delta te_m}$$

Figura 5.5-vii

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + (\frac{1}{4} + \frac{1}{2}) \Delta te_m}$$

Figura 5.5-viii

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + (\frac{1}{2} + \frac{1}{4}) \Delta te_m}$$

Figura 5.5-ix

Figura 5.5-x

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \left(\frac{1}{4} + \frac{1}{4}\right) \Delta te_m}$$

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \left(\frac{3}{4} + \frac{3}{4}\right) \Delta te_m}$$

Figura 5.5-xi

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \left(\frac{1}{8} + \frac{3}{8}\right) \Delta te_m}$$

Figura 5.5-xii

$$ek_{re(I_m)} = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \left(\frac{5}{8} + \frac{7}{8}\right) \Delta te_m}$$

Das expressões correspondentes às situações de 5.5-i até 5.5-xii, observamos que:

$$re(I_m) = \frac{M[te_m] \Delta te_m}{V[te_m^-] \Delta te_m + \sum_{c=1}^{W[I_m]} f_c(I_m) \Delta te_m} = \frac{M[te_m]}{V[te_m^-] + \sum_{c=1}^{W[I_m]} f_c(I_m)}, \quad (5.40)$$

onde:

- 1- c é o índice relacionado com os pontos de censura
- 2- $f_c(I_m)$ representa a fração do comprimento de I_m que corresponde à distância da c -ésima censura deste subintervalo em relação ao ponto inicial te_{m-1} .

Portanto, como $M[te_m]$ deve ser sempre 1, a expressão do esttimador de Kitchin em termos de funções empíricas, nos pontos te_s ($s \geq m$), de "mortes^e" (isto é, a sua "parte discreta"), é a seguinte:

$$Se(te_s) = \exp - \sum_{te_m \leq te_s} \frac{1}{nS_T^n(te_m^-) + \sum_{c=1}^{W[I_m]} f_c(I_m)}. \quad (5.41)$$

Finalmente, ressalte-se que, como comentado no capítulo 3, os estimadores bayesianos (sv) e (sp) podem se transformar em versões modificadas do estimador (km). Conseqüentemente, as estimativas das suas taxas acumuladas $re(I_m)$ tendem a se aproximar das correspondentes estimativas (km).

6. PROPOSTA DE UM NOVO ESTIMADOR F-N-P, ALTERNATIVO AOS ESTIMADORES RES F-N-P ESTUDADOS.

Conforme será visto agora, comparando-se as formas funcionais, estabelecidas nos capítulos 2 e 4, para a parte discreta de todos os estimadores estudados (exceto os bayesianos), serão identificadas duas classes de estimadores.

Dentro de cada uma delas será feita uma análise mais detalhada e comparativa dos fatores componentes da expressão de cada estimador, ressaltando-se as suas principais semelhanças e diferenças. No final deste trabalho, depois de ter-se estabelecido o estimador alternativo, retomar-se-á esta análise, incluindo-se o novo estimador.

Seguindo uma condição básica dos objetivos deste trabalho, será fixada primeiro uma das classes funcionais de estimadores, para definir a forma geral que o estimador a ser proposto deva satisfazer. Nesta classe será escolhido um de seus estimadores, com características a serem reproduzidas no novo estimador e com possibilidades para a implementação de algum tipo de simplificação em sua taxa de "mortes"^e, sem prejuízo na qualidade das estimativas.

Observando-se as ordenações das taxas F-N-P, estabelecidas no capítulo 5, somente com os estimadores da classe funcional escolhida para o estimador alternativo, será definida a "taxa modificada", com base na interpretação geométrica até aqui adotada. Os detalhes destes procedimentos estão expostos nos próximos itens.

6.1. Comparação das formas funcionais na parte discreta dos estimadores F-N-P estudados.

Os estimadores são organizados pela ordem cronológica com que foram divulgados ou passaram a ser utilizados nas aplicações, conforme indicam as publicações^{4,14,42}.

De (4.37), (4.29), (4.20), (5.40) e (2.3), tem-se, respectivamente:

$$^{ea} Se(te_s) = \exp - \sum_{te_m \leq te_s} -\ln \left[1 - \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}} \right], \quad (6.1)$$

$$^{km} Se(te_s) = \exp - \sum_{te_m \leq te_s} -\ln \left[1 - \frac{M[te_m]}{V[te_m^-]} \right], \quad (6.2)$$

$$^{en} Se(te_s) = \exp - \sum_{te_m \leq te_s} \frac{M[te_m]}{V[te_m^-]}, \quad (6.3)$$

$$^{ek} Se(te_s) = \exp - \sum_{te_m \leq te_s} \left[\frac{M[te_m]}{V[te_m^-] + \sum_{c=1} f_c(I_m)} \right], \quad (6.4)$$

$$^{ep} Se(te_s) = \exp - \sum_{te_m \leq te_s} \frac{M[te_m]}{V[te_m^-] + \frac{W[I_m]}{2}}. \quad (6.5)$$

Note-se que as expressões (6.1) e (6.2) e depois (6.3), (6.4) e (6.5) têm a mesma forma. Assim, identificam-se duas classes de estimadores :

Classe 1- composta dos estimadores (en) , (ek) e (ep)

e

Classe 2- composta dos estimadores (ea) , (km) .

Na primeira, as taxas acumuladas em cada subintervalo são, conforme as interpretações geométricas do capítulo 5, simplesmente calculadas por uma razão de áreas, enquanto, na segunda, elas são calculadas pelo logaritmo (de base e) de uma razão de áreas.

Retirando o fator $W[I_m]/2$ de (6.1), obtem-se (6.2) e de (6.5), obtem-se (6.3) e retirando a soma $\delta = \sum [f_c(I_m)]$ de (6.4), obtem-se (6.3). Em (6.4), a retirada da soma δ equivale a considerar todas as distâncias f_c iguais a zero. Isso pode ser interpretado como a consideração de todas as censuras em I_m posicionadas no início deste subintervalo.

Se, em (6.4), imagina-se todas as distâncias f_c iguais a $1/2$ (isto é, todas as censuras posicionadas no meio do subintervalo) obtem-se (6.5) [veja-se, por exemplo, as figuras 5.5 ii e vi]. Note-se também, que, para algumas situações do tipo representado nas figuras 5.5 (como a figura 5.5-vi, por exemplo), a expressão (6.4) se iguala à (6.5).

Para um subintervalo I_m sem censuras, ter-se-á $W[I_m]=0$. Em consequência, os estimadores da Classe 1 reduzem-se ao (6.3), enquanto os da Classe 2 reduzem-se ao (6.2) [veja-se, por exemplo a figura 5.5-i].

6.2. Proposta do novo estimador F-N-P, não paramétrico, alternativo aos apresentados.

No capítulo anterior, mostrou-se [veja-se (5.29)] que em cada ponto de "morte"^e [respeitando-se as restrições impostas em (5.21) e em (5.24)] são sempre válidas as seguintes ordenações:

$$\begin{aligned}
A) \quad & \overset{ek}{re(I_m)} \leq \overset{en}{re(I_m)} < \overset{km}{re(I_m)}, \\
B) \quad & \overset{ep}{re(I_m)} \leq \overset{en}{re(I_m)} < \overset{km}{re(I_m)}, \\
C) \quad & \overset{ep}{re(I_m)} < \overset{ea}{re(I_m)} \leq \overset{km}{re(I_m)}.
\end{aligned} \tag{6.6}$$

Para grandes amostras, as estimativas nestas desigualdades tendem a se tornar muito próximas umas das outras, tendendo a uma estimativa limite da taxa $re(I_m)$. Neste caso torna-se quase indiferente a escolha de uma ou outra dessas estimativas.

Por outro lado, em estudos de seguimentos clínicos, como o de transplantes de rim ilustrado aqui, em geral, as amostras são muito pequenas, tornando as desigualdes (6.6) relevantes. Neste caso, um procedimento inicial, julgado razoável para a definição de um estimador alternativo aos apresentados, é o de procurar "espaços" entre cada duas estimativas [considerando cada uma das três ordenações em (6.6)], a serem ocupados pelas estimativas desejadas. Com isso, garante-se que, com o aumento do tamanho das amostras, essas "novas" estimativas tendem à situação limite quase tão rapidamente quanto as estimativas que a delimitam.

As possibilidades de espaços diminuem a partir do momento em que se fixa o tipo de forma funcional que se deseja para o estimador. Com essa finalidade, e atendendo a condição prevista na primeira parte dos objetivos deste trabalho, escolheu-se definir essa taxa segundo o critério, matematicamente mais simples, utilizado nos estimadores da classe 1. Assim, as 6 (seis) possibilidades em (6.6) reduzem-se às duas:

$$\begin{aligned}
A') \quad & \overset{ek}{re(I_m)} \leq \overset{en}{re(I_m)}, \\
B') \quad & \overset{ep}{re(I_m)} \leq \overset{en}{re(I_m)}.
\end{aligned} \tag{6.7}$$

Para maior simplicidade, optou-se por definir a taxa instantânea alternativa no ponto te_m pela média aritmética das taxas instantâneas neste ponto, segundo os estimadores dos extremos no espaço "vazio" escolhido. Como consequência a taxa acumulada alternativa no correspondente subintervalo é, também, a média aritmética das respectivas taxas acumuladas.

A "suavização" na estimativa da curva de sobrevivência, gerada pela continuidade do estimador de Kitchin, é uma característica que se deseja ter no novo estimador, para refletir uma das realidades da curva de sobrevivência teórica a ser estimada. Além disso, este estimador é claramente, dentre os três na classe 1, o que oferece maior potencial de simplificação em sua taxa. Portanto, fixa-se o estimador de Kitchin como modelo a ser simplificado para dar origem ao novo estimador, que daqui em diante será chamado "estimador modificado de Kitchin" [e identificado pela sigla (mk)].

Em (6.7), observa-se que tanto a taxa de Kitchin como a taxa padrão são menores ou iguais à taxa de Nelson (isto é, as correspondentes curvas de sobrevivência de Kitchin e padrão em um mesmo ponto de "morte^e" são mais altas que a de Nelson). Mas a maior simplicidade funcional do estimador da taxa padrão e a sua maior semelhança com a forma funcional da taxa de Nelson facilita a manipulação algébrica no estabelecimento da expressão de uma taxa média entre essas duas. Assim, decidiu-se definir a "taxa modificada" de "morte^e" como a média aritmética das taxas padrão e de Nelson.

Então, segundo a interpretação por áreas, adotada no capítulo 5, a taxa modificada instantânea, no ponto te_m , é definida por:

$${}^{mk}re(te_m) = \frac{{}^{en}re(te_m) + {}^{ep}re(te_m)}{2}, \quad (6.8)$$

isto é:

$$mk \quad re(te_m) = \frac{\frac{1}{A_{en}} + \frac{1}{\bar{A}}}{2} \quad (6.9)$$

ou

$$mk \quad re(te_m) = \frac{\text{INVERSO DA MÉDIA HARMÔNICA DAS ÁREAS } \bar{A} \text{ E } A_{en}}{2} \quad (6.10)$$

Em consequência, a correspondente taxa acumulada no subintervalo I_m fica definida por:

$$mk \quad re(I_m) = \frac{A_b}{\text{MÉDIA HARMÔNICA DE } \bar{A} \text{ E } A_{en}} \quad (6.11)$$

Além disso, por (5.2), (5.3) e (5.14):

$$\begin{aligned} mk \quad re(I_m) &= \frac{M[te_m] \Delta te_m}{2} \left[\frac{1}{V(te_m^-) \Delta te_m} + \frac{1}{(V(te_m^-) + W[I_m]/2) \Delta te_m} \right] \\ &= \frac{M[te_m]}{2V(te_m^-)} \left[\frac{2V(te_m^-) + W[I_m]/2}{V(te_m^-) + W[I_m]/2} \right] = \frac{M[te_m]}{V(te_m^-)} \left[\frac{V(te_m^-) + W[I_m]/4}{V(te_m^-) + W[I_m]/2} \right] \\ &= \frac{\frac{M[te_m]}{V(te_m^-)}}{\frac{V(te_m^-) + W[I_m]/2}{V(te_m^-) + W[I_m]/4}} = \frac{\frac{M[te_m]}{V(te_m^-)}}{1 + \frac{W[I_m]/4}{V(te_m^-) + W[I_m]/4}} \end{aligned}$$

$$= \frac{M[te_m]}{V(te_m^-) + \frac{W[I_m]/4}{1 + \frac{W[I_m]/4}{V(te_m^-)}}} = \frac{M[te_m]}{V(te_m^-) + \frac{W[I_m]}{4 + \frac{W[I_m]}{V(te_m^-)}}} \quad (6.12)$$

Propõe-se, então, uma taxa modificada com a forma funcional das taxas dos estimadores da classe 1.

De maneira análoga ao que foi feito de (6.1) até (6.5), es-creve-se a parte discreta do estimador modificado de Kitchin como:

$$Se(te_s) = \exp - \sum_{te_m \leq te_s} \left[\frac{M[te_m]}{V(te_m^-) + \frac{W[I_m]}{4 + \frac{W[I_m]}{V(te_m^-)}}} \right] \quad (6.13)$$

Da mesma forma como no estimador de Kitchin, este estima-dor é contínuo e, portanto, $M[te_m]$ deve ser sempre igual a 1.

6.3. Uma análise dos fatores que diferenciam os estimado- res com a forma funcional da classe 1.

Note-se que, no estimador padrão, o fator $W[I_m]/2$ depende somente do número de censuras dentro do subintervalo, podendo ser in-terpretado como uma espécie de "reposição" de metade das censuras que já foram contabilizadas em $V(te_m^-)$. Portanto, em cada I_m , sempre se atribui, nessas estimativas, um "efeito" correspondente à metade das censuras ocorridas naquele subintervalo, independentemente deste se situar no início, meio ou final do intervalo total de tempo de so-brevivência.

Dentro desta interpretação, no estimador de Nelson não ocorre tal "reposição", pois $W(I_m)/2=0$.

No estimador de Kitchin, o fator correspondente a esse efeito, nos pontos de "morte"^e, é a soma $\delta = \sum [f_c(I_m)]$, de cálculo mais sofisticado que, além de depender do número de censuras, também depende da posição delas no subintervalo.

No estimador modificado esse "efeito" é medido por:

$$\chi = \frac{W(I_m)}{4 + \tau}, \quad (6.14)$$

onde

$$\tau = W(I_m)/V(te_m^-). \quad (6.15)$$

O seu cálculo é menos complicado que o do fator δ de Kitchin (pois só exige contagens), dependendo tanto do número de censuras no subintervalo, como do número ainda em risco antes do seu extremo superior te_m^- . Com o passar do tempo $V(te_m^-)$, vai diminuindo, tornando, deste modo, o fator χ também dependente do tempo.

6.4. Cálculo do estimador (mk), para os dados de transplante de rim, doador vivo, esquema A.

Na figura 6.1, a seguir, apresenta-se uma comparação gráfica das estimativas (ek) e (mk) da curva de sobrevivência. Na tabela A11, no Anexo-A, apresentam-se os detalhes dos cálculos e na tabela 6.1, no final deste capítulo, comparam-se as estimativas (ek) e (mk), das probabilidades de sobrevivência, e suas médias e medianas.

Conforme observa-se na figura 6.1, as estimativas (mk) não

diferem muito das correspondentes estimativas (ek). Deve-se confirmar isto mediante a realização de simulações computacionais.

Figura 6.1 Comparação gráfica entre as estimativas (ek) e (mk), da curva de sobrevivência, para os dados de transplante de rim, doador vivo, esquema A.

Tabela 6.1 Estimativas (ek) e (mk) da curva de sobrevivência e da média e mediana, para os dados de transplante de rim, doador vivo, esquema A.

t_s	$S(t_s)^{(ek)}$	sub-área _s	$S(t_s)^{(mk)}$	sub-área _s
0	1,0000		1,0000	
40	0,9608	39,211	0,9608	39,211
43	0,9216	2,823	0,9216	2,823
89	0,8824	41,526	0,8824	41,526
201	0,8432	96,649	0,8432	96,649
215	0,8039	11,528	0,8039	11,528
308	0,7647	73,057	0,7647	73,057
396	0,7255	65,553	0,7255	65,553
512	0,6863	81,772	0,6863	81,772
542	0,6471	19,984	0,6471	19,984
956*	0,6219		0,6207	
970*	0,6211		0,6198	
1227	0,6060	428,446	0,6039	428,446
1272*	0,5941		0,5912	
1343*	0,5757		0,5716	
1362*	0,5709		0,5665	
1424	0,5555	114,797	0,5500	113,302
1426*	0,5543		0,5487	
1434*	0,5494		0,5436	
1483*	0,5201		0,5132	
1518*	0,5001		0,4926	
1570	0,4719	74,932	0,4634	73,669
1588*	0,4580		0,4491	
1609*	0,4436		0,4330	
1713*	0,3723		0,3615	
1973	0,2419	137,671	0,2301	135,188
MÉDIA		1187,9 dias		1182,7 dias
MEDIANA		1519,8 dias		1503,5 dias

7. SIMULAÇÃO DOS DADOS E COMPARAÇÃO DO DESEMPENHO DOS ESTIMADORES.

Para comparar o desempenho dos estimadores estudados e do proposto, utilizou-se simulação estocástica, em micro-computador, de amostras com dados censurados, semelhantes à do exemplo de transplantes de rim.

Para isso, elaborou-se um programa computacional, que passou a ser denominado de "APOIO", constituído de sub-programas interligados, funcionando interativamente, dispondo de algoritmos para modelar tanto as condições teóricas básicas assumidas nas definições dos estimadores (como independência entre as variáveis aleatórias T_e e T_o , definição das suas distribuições e exigência de ausência de empates) como para executar as etapas exigidas nos procedimentos de cálculo e comparação das estimativas.

Codificou-se os sub-programas em linguagem Pascal (Turbo Pascal, versão 5.0 e Borland Pascal, versão 7.0) e utilizou-se o próprio gerador de números aleatórios implementado neste "software".

Testes estatísticos preliminares (como Qui-quadrado de Kolmogorov e Smirnov) e análises gráficas indicaram uma razoável confiabilidade nas sequências numéricas geradas.

7.1. Esquema de geração das amostras.

Para cada amostra, foram geradas simultaneamente três sequências independentes de números aleatórios com distribuição uniforme no intervalo (0,1), isto é:

$$\begin{array}{ccc} \{x_n\} \sim UC(0,1) & , & \{y_n\} \sim UC(0,1) \quad \text{e} \quad \{z_n\} \sim UC(0,1). \\ \downarrow & & \downarrow \\ x_1 & & y_1 \\ x_2 & & y_2 \\ \vdots & & \vdots \\ x_n & & y_n \end{array} \quad \begin{array}{c} \downarrow \\ z_1 \\ z_2 \\ \vdots \\ z_n \end{array}$$

A primeira e a segunda para produzirem, após uma transfor-

mação conveniente, n tempos t_e (em dias) e n tempos t_o (em dias), respectivamente. A terceira foi planejada para selecionar os valores de cada uma das duas primeiras para comporem a amostra desejada de n tempos, de acordo com uma proporção π fixada de censuras.

O critério de decisão adotado foi o seguinte:

$$\text{valor } u_i \text{ para compor a amostra} = \begin{cases} y_i, & \text{se } z_i \in (0; \pi] \\ x_i, & \text{se } z_i \in (\pi; 1) \end{cases}, \quad i=1, 2, \dots, n. \quad (7.1)$$

Para a transformação desses valores u_i em tempos (dados em dias) seguindo distribuições teóricas pré-determinadas F_e e F_o , com funções de sobrevivência S_e e S_o [(Exponencial e Exponencial), (Exponencial e Weibull), (Weibull e Exponencial) e (Weibull e Weibull)] foram usadas as relações inversas dessas funções de sobrevivência. Isto é:

- para a distribuição exponencial de parâmetro λ ,

$$t_i = -\frac{1}{\lambda} \ln(u_i) \quad (7.2)$$

e - para a distribuição Weibull de parâmetros α e β ,

$$t_i = \frac{1}{\alpha} [-\ln(u_i)]^{1/\beta}. \quad (7.3)$$

Neste programa foi implementado um mecanismo de comparação dos valores, durante a própria geração dos tempos, permitindo a identificação e substituição dos valores empatados por novos valores gerados. Um procedimento semelhante a este assegurou, para cada amostra, o cumprimento exato da quantidade fixada de censuras.

Em cada amostra, os dados gerados foram arquivados em colunas onde os tempos de mortes t_e e de censuras t_o foram identificados em uma coluna paralela, respectivamente, pelos dígitos 1 e 0.

Conforme o esquema acima, em cada processamento foi gerada uma quantidade N de amostras de mesmo tamanho n , com os tempos de

mortes e de censuras gerados, respectivamente, das distribuições teóricas $F_e=1-S_e$ e $F_o=1-S_o$, com uma taxa fixada π de censuras. Assim, foram processadas configurações do tipo:

$$[N;n;S_e;S_o;\pi]. \quad (7.4)$$

7.2. Cálculo do estimador bayesiano de Salinas e Pereira usando os dados simulados.

Da mesma forma como foi feito para o exemplo de transplantes de rim, em cada amostra de uma mesma configuração, adotou-se a priori não informativa com parâmetros:

$$a_1(t) = a_2(t) = 1 - \frac{t}{m} \quad (7.5)$$

com m sendo o tempo máximo gerado em cada amostra.

Também, usou-se $\Delta=1$, para simplificar (3.21), no cálculo numérico das integrais pelo método de Runge e Kutta de ordem 4.

7.3. Cálculo do desempenho dos estimadores usando os dados simulados.

Em cada amostra de uma dada configuração $[N;n;S_e;S_o;\pi]$, calculou-se a curva básica de sobrevivência S_e (sem a influência das censuras). Também, para cada uma dessas N amostras e cada um dos quatro estimadores tratados, calculou-se as N curvas de sobrevivência estimadas.

Fixando uma sequência de tempos, determinou-se as correspondentes probabilidades de sobrevivência, em cada uma dessas curvas estimadas para, mediante a maneira que será comentada em seguida, compará-las com respeito à curva básica calculada de sobrevivência e avaliar o desempenho do seu estimador.

Aqui, definiu-se essa sequência de pontos, procurando os valores L_1 e L_2 que fossem superados aproximadamente por 0,1% dos valores t_e e t_o , das respectivas distribuições teóricas. Isto é:

$$\begin{aligned} Se(L_1) &\cong 0,001 \\ Se(L_2) &\cong 0,001 \end{aligned} \quad (7.6)$$

e considerou-se o intervalo $(0, L)$,

$$\text{com} \quad L = \max\{L_1, L_2\}. \quad (7.7)$$

Em seguida, fixou-se d =número de pontos a serem inseridos no intervalo $(0, L)$ e calculou-se:

$$h = \frac{L}{d}, \quad (7.8)$$

para funcionar como incrementos iguais a serem adicionados a partir do ponto 0 (zero) e, desta forma, determinar cada ponto dentro do intervalo $(0, L)$. Isto é, os pontos ficaram definidos por:

$$w_i = ih, \quad \text{com } i = 1, 2, \dots, d-1, \quad (7.9)$$

fornecendo a sequência desejada.

Para cada um dos estimadores $Se(.)$ determinou-se, respectivamente, o seu vício (EM) e erro quadrático médio (EQM), em um ponto fixado w_i , dessa sequência, simplesmente calculando as médias dos desvios

$$Se(w_i) - Se(w_i) \quad (7.10)$$

$$e \quad [Se(w_i) - Se(w_i)]^2, \quad (7.11)$$

levando-se em conta todas as N amostras na configuração, ou seja:

$$EMC(w_i) = \frac{1}{N} \sum_{j=1}^N [Se(w_i) - Se(w_i)]_j \quad (7.12)$$

$$EQM(w_i) = \frac{1}{N} \sum_{j=1}^N ([Se(w_i) - Se(w_i)]^2)_j, \quad (7.13)$$

7.4. Resultados observados nos dados simulados.

Os valores registrados em cada uma das tabelas, de 7.1 a 7.16, neste capítulo, foram calculados utilizando configurações com $N=500$ amostras, de tamanhos $n=10$ ou $n=50$, de dados simulados segundo os métodos descritos acima, com distribuições Fe e Fo das famílias exponencial e Weibull, conforme valores escolhidos dos parâmetros e com as seguintes proporções de censuras $\pi=0$ ou $\pi=0,20$ ou $\pi=0,40$.

O número de amostras ($N=500$) foi usado pela necessidade de se adequar o volume e complexidade dos cálculos envolvidos às condições de velocidade de processamento e capacidades de memória do equipamento computacional disponível durante a fase de elaboração do programa. Na fase final, este número foi mantido porque não houve diferenças nos resultados de algumas simulações com $N = 1000$ amostras e também, para permitir o maior número de testes de distintas configurações, dentro de tempo programado para as simulações.

As simulações definitivas foram realizadas em equipamento computacional mais avançado, com micro-processador "Pentium" de 133 MHz, disco rígido de 1,7 GB e memória básica de 16 MB.

As tabelas têm aqui a finalidade somente de comparação dos estimadores em pequenas amostras, sem a pretensão de fornecer estimativas de alta acurácia do vício e erro quadrático médio. Mas elas são significativas mesmo quando algumas amostras simuladas, dentro de cada configuração, se desviarem das distribuições que elas deveriam representar.

A seguir, para as configurações das tabelas 7.1 a 7.16, foram construídas as figuras 7.1 a 7.16, para descrever o comportamento dos estimadores km , ek , mk e sp , com relação ao erro quadrático médio (EQM) e viés (EM). Quando $\pi=0$ as estimativas (kt) e (mk) coincidem e portanto os correspondentes EQM e EM também coincidem (veja-se as figuras 7.1, 7.2, 7.13 e 7.14).

Tabela 7.1 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;exp($\lambda=0.0015$);exp($\lambda=0.0015$); $\pi=0$].

ERRO QUADRATICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.003719	0.003705	0.003705	0.003658
460	0.004614	0.004635	0.004635	0.004692
690	0.004655	0.004685	0.004685	0.004820
920	0.003909	0.003989	0.003989	0.004206
1150	0.003218	0.003345	0.003345	0.003629
1380	0.002123	0.002232	0.002232	0.002502
1610	0.001612	0.001707	0.001707	0.001977
1840	0.001167	0.001281	0.001281	0.001543
2070	0.000875	0.000984	0.000984	0.001236
2300	0.000613	0.000700	0.000700	0.000927
2530	0.000396	0.000488	0.000488	0.000674
2760	0.000297	0.000395	0.000395	0.000556
2990	0.000208	0.000310	0.000310	0.000425
3220	0.000157	0.000260	0.000260	0.000332
3450	0.000112	0.000218	0.000218	0.000251
3680	0.000071	0.000179	0.000179	0.000172
3910	0.000053	0.000164	0.000164	0.000132
4140	0.000033	0.000145	0.000145	0.000085
4370	0.000021	0.000134	0.000134	0.000059
VIÉS				
230	0.008900	0.011712	0.011712	0.016858
460	0.008904	0.013775	0.013775	0.021739
690	0.006494	0.012854	0.012854	0.022058
920	0.006941	0.014343	0.014343	0.023514
1150	0.007747	0.015897	0.015897	0.024133
1380	0.004534	0.013273	0.013273	0.020180
1610	0.002713	0.011897	0.011897	0.017023
1840	0.002988	0.012512	0.012512	0.015782
2070	0.002175	0.012001	0.012001	0.013299
2300	0.000534	0.010644	0.010644	0.010020
2530	0.000277	0.010602	0.010602	0.008200
2760	0.000277	0.010800	0.010800	0.006675
2990	0.000083	0.010756	0.010756	0.005059
3220	-0.000067	0.010732	0.010732	0.003676
3450	-0.000176	0.010716	0.010716	0.002670
3680	-0.000326	0.010634	0.010634	0.001714
3910	-0.000277	0.010731	0.010731	0.001243
4140	-0.000369	0.010677	0.010677	0.000616
4370	-0.000383	0.010689	0.010689	0.000317

Figura 7.1 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0.0015); \exp(\lambda=0.0015); \pi=0]$.

Tabela 7.2 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp($\lambda=0.0015$);exp($\lambda=0.0015$); $\pi=0$].

ERRO QUADRATICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.086581	0.084574	0.084574	0.081922
460	0.026023	0.024194	0.024194	0.023070
690	0.026372	0.024385	0.024385	0.024726
920	0.020356	0.019157	0.019157	0.022075
1150	0.014839	0.014575	0.014575	0.018637
1380	0.010928	0.011431	0.011431	0.016298
1610	0.008246	0.008963	0.008963	0.013704
1840	0.005891	0.007166	0.007166	0.010762
2070	0.004471	0.005982	0.005982	0.008614
2300	0.002954	0.004885	0.004885	0.006313
2530	0.001893	0.003986	0.003986	0.004425
2760	0.001327	0.003606	0.003606	0.003334
2990	0.000700	0.002977	0.002977	0.001847
3220	0.000450	0.002839	0.002839	0.001266
3450	0.000298	0.002800	0.002800	0.000931
3680	0.000274	0.002936	0.002936	0.000850
3910	0.000159	0.002846	0.002846	0.000495
4140	0.000081	0.002789	0.002789	0.000293
4370	0.000041	0.002779	0.002779	0.000164
VIÉS				
230	-0.127020	-0.113578	-0.113578	-0.096608
460	-0.007976	0.016250	0.016250	0.037685
690	-0.005626	0.026421	0.026421	0.044174
920	-0.005779	0.031805	0.031805	0.042571
1150	-0.003773	0.037799	0.037799	0.040219
1380	-0.001786	0.042827	0.042827	0.036340
1610	-0.004767	0.042480	0.042480	0.025900
1840	-0.003292	0.045633	0.045633	0.020910
2070	-0.004025	0.046320	0.046320	0.014035
2300	-0.002546	0.048631	0.048631	0.011497
2530	-0.003083	0.048826	0.048826	0.007266
2760	-0.002523	0.049856	0.049856	0.005315
2990	-0.003877	0.048972	0.048972	0.000711
3220	-0.003387	0.049690	0.049690	-0.000191
3450	-0.002656	0.050551	0.050551	-0.000254
3680	-0.001206	0.052018	0.052018	0.000981
3910	-0.001237	0.052085	0.052085	0.000056
4140	-0.001209	0.052178	0.052178	-0.000387
4370	-0.001023	0.052396	0.052396	-0.000509

Figura 7.2 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0.0015); \exp(\lambda=0.0015); \pi=0]$.

Tabela 7.3 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;exp($\lambda=0.0015$);exp($\lambda=0.0015$); $\pi=0.40$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.013774	0.013594	0.013437	0.014516
460	0.030717	0.031023	0.030486	0.032982
690	0.039278	0.040754	0.039833	0.042884
920	0.039005	0.041258	0.040004	0.043433
1150	0.037330	0.040673	0.039074	0.042008
1380	0.032512	0.036125	0.034214	0.037063
1610	0.028013	0.031966	0.029783	0.032031
1840	0.023188	0.027162	0.024737	0.026408
2070	0.017874	0.022833	0.020285	0.020572
2300	0.014358	0.019490	0.017000	0.016434
2530	0.012406	0.017776	0.015272	0.014082
2760	0.010064	0.015451	0.012972	0.010882
2990	0.008713	0.014629	0.012170	0.008972
3220	0.007830	0.013819	0.011354	0.007667
3450	0.006428	0.012853	0.010438	0.005816
3680	0.005623	0.012309	0.009924	0.004643
3910	0.005067	0.011923	0.009552	0.003626
4140	0.004923	0.011823	0.009467	0.003365
4370	0.004903	0.011823	0.009468	0.003224
VIÉS				
230	0.104352	0.104138	0.103370	0.108909
460	0.162370	0.164121	0.162425	0.170289
690	0.184711	0.189580	0.187029	0.195429
920	0.182572	0.189861	0.186426	0.195662
1150	0.177416	0.187739	0.183265	0.191664
1380	0.165037	0.177373	0.171851	0.179589
1610	0.150503	0.165684	0.158972	0.164609
1840	0.134047	0.152003	0.144015	0.146615
2070	0.114174	0.138714	0.129764	0.125746
2300	0.097975	0.127188	0.118023	0.107468
2530	0.085176	0.119941	0.110514	0.092596
2760	0.071122	0.110334	0.100686	0.075068
2990	0.064967	0.107828	0.098268	0.066226
3220	0.060372	0.105041	0.095393	0.059611
3450	0.052478	0.101252	0.091551	0.049423
3680	0.047631	0.098972	0.089280	0.042811
3910	0.043952	0.097121	0.087437	0.037229
4140	0.042756	0.096547	0.086923	0.035449
4370	0.042417	0.096382	0.086793	0.034474

Figura 7.3 Comportamento dos estimadores km , ek , mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0,0015); \exp(\lambda=0,0015); \pi=0,40]$.

Tabela 7.4 Desempenho dos estimadores com os dados simulados na configuração $[N=500; n=10; \exp(\lambda=0.0015); \exp(\lambda=0.0015); \pi=0.40]$.

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.068647	0.066118	0.065959	0.068318
460	0.043871	0.043669	0.042255	0.046809
690	0.058535	0.061052	0.057647	0.064927
920	0.062431	0.069744	0.064546	0.071340
1150	0.060606	0.071216	0.064562	0.069305
1380	0.053123	0.067918	0.060644	0.061111
1610	0.050483	0.069344	0.061460	0.056696
1840	0.044897	0.065972	0.057841	0.047200
2070	0.039287	0.064412	0.055996	0.039669
2300	0.035778	0.062032	0.053415	0.033900
2530	0.031221	0.059880	0.051363	0.026810
2760	0.028376	0.059138	0.050764	0.021701
2990	0.026075	0.058142	0.049814	0.017742
3220	0.025364	0.058004	0.049667	0.015963
3450	0.024874	0.058726	0.050320	0.015499
3680	0.024582	0.059041	0.050592	0.014802
3910	0.024305	0.058967	0.050533	0.014352
4140	0.023619	0.057947	0.049607	0.012471
4370	0.023103	0.057448	0.049111	0.011422
VIÉS				
230	-0.006811	-0.005475	-0.007587	0.011487
460	0.133796	0.147885	0.141711	0.160733
690	0.179647	0.205335	0.195110	0.208113
920	0.181161	0.223936	0.210553	0.209385
1150	0.175753	0.230570	0.214714	0.197914
1380	0.150868	0.226599	0.209724	0.168045
1610	0.145976	0.231910	0.214540	0.156342
1840	0.135005	0.228616	0.211234	0.135493
2070	0.124572	0.228962	0.211339	0.120418
2300	0.116231	0.226240	0.208391	0.107829
2530	0.105471	0.223725	0.206033	0.091888
2760	0.099284	0.223112	0.205681	0.080546
2990	0.095415	0.222280	0.205009	0.074141
3220	0.093776	0.222112	0.204936	0.070610
3450	0.093843	0.223824	0.206609	0.070768
3680	0.094156	0.224580	0.207381	0.070217
3910	0.093671	0.224522	0.207406	0.069514
4140	0.091357	0.222667	0.205669	0.065016
4370	0.089542	0.221473	0.204499	0.062219

Figura 7.4 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0; 0.0015), \exp(\lambda=0.0015); \pi=0.40]$.

Tabela 7.5 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;exp($\lambda=0.007$);exp($\lambda=0.03$); $\pi=0.20$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
50	0.006325	0.006304	0.006225	0.006920
100	0.008645	0.009247	0.009101	0.010035
150	0.008345	0.009261	0.009120	0.010065
200	0.006548	0.007507	0.007410	0.008214
250	0.005157	0.005987	0.005919	0.006602
300	0.003742	0.004429	0.004383	0.004938
350	0.002830	0.003419	0.003386	0.003875
400	0.001936	0.002440	0.002417	0.002853
450	0.001396	0.001746	0.001732	0.002147
500	0.001009	0.001316	0.001306	0.001664
550	0.000657	0.000913	0.000905	0.001177
600	0.000461	0.000705	0.000699	0.000894
650	0.000349	0.000596	0.000592	0.000726
700	0.000236	0.000480	0.000477	0.000520
750	0.000151	0.000377	0.000375	0.000363
800	0.000122	0.000364	0.000361	0.000311
850	0.000081	0.000312	0.000310	0.000208
900	0.000045	0.000272	0.000270	0.000126
950	0.000028	0.000252	0.000251	0.000080
VIÉS				
50	0.059974	0.060856	0.060204	0.066314
100	0.065385	0.070638	0.069617	0.077582
150	0.057464	0.066062	0.065078	0.073738
200	0.045232	0.055766	0.054962	0.063708
250	0.034591	0.046214	0.045594	0.053687
300	0.024871	0.037301	0.036851	0.043334
350	0.019041	0.031984	0.031648	0.036157
400	0.014387	0.027760	0.027508	0.029711
450	0.007942	0.021771	0.021589	0.020868
500	0.005651	0.019815	0.019672	0.016545
550	0.002967	0.017434	0.017321	0.011724
600	0.002044	0.016734	0.016642	0.009038
650	0.001758	0.016608	0.016530	0.007496
700	0.001232	0.016211	0.016140	0.005557
750	0.000348	0.015448	0.015387	0.003488
800	0.000734	0.015882	0.015825	0.003424
850	0.000199	0.015416	0.015362	0.002005
900	-0.000078	0.015181	0.015130	0.001127
950	-0.000247	0.015045	0.014995	0.000547

Figura 7.5 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0.007); \exp(\lambda=0.03); \pi=0.20]$.

Tabela 7.6 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp($\lambda=0.007$);exp($\lambda=0.03$); $\pi=0.20$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
50	0.035963	0.033145	0.033048	0.033531
100	0.029125	0.027852	0.027405	0.028940
150	0.031733	0.031915	0.031366	0.034984
200	0.036877	0.028091	0.027546	0.033100
250	0.021005	0.023195	0.022657	0.029541
300	0.015866	0.019050	0.018584	0.024991
350	0.012124	0.016281	0.015900	0.021045
400	0.008610	0.013246	0.012932	0.016527
450	0.006735	0.011635	0.011370	0.014105
500	0.005324	0.010641	0.010388	0.011633
550	0.004009	0.009513	0.009307	0.008873
600	0.002762	0.008206	0.008017	0.006065
650	0.001940	0.007364	0.007177	0.004279
700	0.001400	0.006828	0.006655	0.003091
750	0.000605	0.005939	0.005791	0.001689
800	0.000449	0.005872	0.005727	0.001384
850	0.000181	0.005539	0.005399	0.000519
900	0.000141	0.005584	0.005442	0.000410
950	0.000141	0.005664	0.005521	0.000411
VIÉS				
50	-0.000352	0.008178	0.005399	0.026630
100	0.043503	0.071820	0.067916	0.086048
150	0.034480	0.076669	0.073120	0.083348
200	0.026004	0.077655	0.074752	0.074807
250	0.016959	0.075786	0.073415	0.062444
300	0.014501	0.077037	0.075060	0.054757
350	0.012601	0.078566	0.076973	0.046061
400	0.010178	0.078106	0.076722	0.038285
450	0.007950	0.077524	0.076321	0.031350
500	0.008181	0.078728	0.077606	0.027178
550	0.007382	0.078650	0.077629	0.021854
600	0.004676	0.076659	0.075686	0.014630
650	0.002824	0.075327	0.074373	0.009760
700	0.001661	0.074623	0.073725	0.006753
750	-0.000651	0.072832	0.072021	0.002880
800	-0.000388	0.073196	0.072397	0.002679
850	-0.001263	0.072492	0.071704	0.000281
900	-0.000771	0.073012	0.072222	0.000579
950	-0.000229	0.073554	0.072764	0.001121

Figura 7.6 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0.007); \exp(\lambda=0.03); \pi=0.20]$.

Tabela 7.7 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;exp(λ =0.007);exp(λ =0.03); π =0.40].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
50	0.013716	0.013234	0.012918	0.014455
100	0.022295	0.023490	0.022776	0.024575
150	0.020031	0.022341	0.021649	0.023233
200	0.015437	0.018160	0.017616	0.018767
250	0.010457	0.012878	0.012522	0.013451
300	0.007231	0.009298	0.009051	0.009859
350	0.005412	0.007079	0.006915	0.007661
400	0.003528	0.004796	0.004690	0.005383
450	0.002566	0.003579	0.003509	0.004166
500	0.001986	0.002864	0.002810	0.003358
550	0.001353	0.002142	0.002102	0.002466
600	0.000922	0.001622	0.001593	0.001769
650	0.000636	0.001281	0.001258	0.001317
700	0.000430	0.001019	0.001001	0.000927
750	0.000256	0.000809	0.000795	0.000594
800	0.000163	0.000700	0.000688	0.000414
850	0.000114	0.000643	0.000633	0.000295
900	0.000049	0.000550	0.000541	0.000140
950	0.000026	0.000527	0.000518	0.000074
VIÉS				
50	0.104250	0.103167	0.101629	0.108854
100	0.128403	0.133859	0.131142	0.139090
150	0.111790	0.123538	0.120752	0.128717
200	0.088510	0.104348	0.101957	0.109629
250	0.065227	0.083218	0.081371	0.088049
300	0.048455	0.067658	0.066261	0.071255
350	0.035933	0.055837	0.054802	0.057056
400	0.023910	0.044467	0.043696	0.042340
450	0.016382	0.037531	0.036953	0.032236
500	0.012145	0.033746	0.033285	0.025305
550	0.009072	0.031012	0.030636	0.019912
600	0.006148	0.028387	0.028071	0.014531
650	0.004199	0.026665	0.026390	0.010744
700	0.002398	0.025053	0.024812	0.007147
750	0.001187	0.023971	0.023752	0.004558
800	0.000506	0.023383	0.023178	0.003001
850	0.000192	0.023132	0.022937	0.002000
900	-0.000612	0.022397	0.022213	0.000364
950	-0.000660	0.022375	0.022196	-0.000060

Figura 7.7 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0.007); \exp(\lambda=0.03); \pi=0.40]$.

Tabela 7.8 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp(λ =0.007);exp(λ =0.03); π =0.40].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
50	0.027713	0.021696	0.020954	0.024756
100	0.044325	0.042147	0.040311	0.043586
150	0.045296	0.048143	0.045998	0.051477
200	0.038748	0.043955	0.042152	0.048846
250	0.030541	0.036738	0.035250	0.042241
300	0.023068	0.029779	0.028620	0.035360
350	0.017862	0.026256	0.025268	0.029979
400	0.012234	0.021273	0.020476	0.022475
450	0.010229	0.020619	0.019848	0.019378
500	0.007525	0.018762	0.018025	0.015449
550	0.005900	0.017533	0.016867	0.012713
600	0.004322	0.016066	0.015443	0.009412
650	0.002915	0.014726	0.014112	0.007052
700	0.002258	0.014232	0.013641	0.005563
750	0.001642	0.013537	0.012965	0.004044
800	0.001319	0.013254	0.012692	0.003148
850	0.001066	0.013008	0.012460	0.002503
900	0.000558	0.012262	0.011729	0.001219
950	0.000421	0.012107	0.011596	0.000714
VIÉS				
50	0.080825	0.081205	0.074421	0.097854
100	0.083659	0.121215	0.112374	0.119984
150	0.074987	0.134373	0.126426	0.120792
200	0.050874	0.125755	0.119558	0.098296
250	0.033129	0.118498	0.113568	0.076049
300	0.017322	0.109484	0.105477	0.054935
350	0.015030	0.110738	0.107362	0.047562
400	0.008100	0.107142	0.104208	0.034263
450	0.011243	0.111495	0.108743	0.032928
500	0.010462	0.112302	0.109665	0.028636
550	0.010027	0.112619	0.110129	0.025193
600	0.007739	0.111277	0.108902	0.018929
650	0.005218	0.109790	0.107523	0.014508
700	0.005023	0.109867	0.107644	0.012510
750	0.003753	0.108909	0.106724	0.009352
800	0.003431	0.108748	0.106586	0.007837
850	0.003057	0.108491	0.106356	0.006620
900	0.001337	0.107012	0.104898	0.003437
950	0.001129	0.106845	0.104762	0.002556

Figura 7.8 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0.007); \exp(\lambda=0.03); \pi=0.40]$.

Tabela 7.9 Desempenho dos estimadores com os dados simulados na configuração [N=500; n=50; exp($\lambda=0.0015$); Weib($\alpha=0.0015, \beta=1.3$); $\pi=0.20$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.006568	0.006625	0.006599	0.007017
460	0.011457	0.011798	0.011701	0.012901
690	0.013402	0.014277	0.014112	0.015544
920	0.012225	0.013431	0.013213	0.014698
1150	0.010636	0.012038	0.011790	0.013137
1380	0.008220	0.009635	0.009389	0.010367
1610	0.005592	0.006918	0.006690	0.007403
1840	0.003735	0.004811	0.004623	0.005154
2070	0.002823	0.003721	0.003545	0.003995
2300	0.002091	0.003003	0.002851	0.003094
2530	0.001541	0.002391	0.002253	0.002358
2760	0.001097	0.001913	0.001783	0.001731
2990	0.000778	0.001568	0.001455	0.001285
3220	0.000564	0.001341	0.001233	0.000958
3450	0.000396	0.001174	0.001064	0.000697
3680	0.000321	0.001091	0.000987	0.000555
3910	0.000258	0.001013	0.000911	0.000412
4140	0.000195	0.000950	0.000850	0.000288
4370	0.000184	0.000953	0.000853	0.000256
VIÉS				
230	0.055205	0.056611	0.056382	0.061335
460	0.082876	0.086082	0.085525	0.093052
690	0.090734	0.096326	0.095443	0.103741
920	0.086534	0.094192	0.092997	0.101377
1150	0.080218	0.089789	0.088338	0.095838
1380	0.064912	0.076768	0.075094	0.080781
1610	0.049849	0.063796	0.062117	0.065386
1840	0.035291	0.051074	0.049318	0.049716
2070	0.026962	0.043954	0.042221	0.039701
2300	0.019486	0.039284	0.037711	0.030416
2530	0.015351	0.036127	0.034638	0.024353
2760	0.011923	0.033482	0.032048	0.018813
2990	0.008266	0.030911	0.029625	0.013615
3220	0.006339	0.029391	0.028169	0.010442
3450	0.004918	0.028372	0.027180	0.007970
3680	0.004451	0.028006	0.026870	0.006790
3910	0.003563	0.027288	0.026186	0.004997
4140	0.002939	0.026795	0.025709	0.003748
4370	0.003137	0.027011	0.025930	0.003723

Figura 7.9 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0.0015); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0.20]$.

Tabela 7.10 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp($\lambda=0.0015$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.20$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.082290	0.080757	0.080694	0.080533
460	0.033639	0.033325	0.033052	0.034475
690	0.032043	0.033382	0.032664	0.035919
920	0.029296	0.032417	0.031469	0.035818
1150	0.024965	0.028294	0.027189	0.032935
1380	0.020496	0.025770	0.024549	0.029118
1610	0.016609	0.022911	0.021644	0.024728
1840	0.013521	0.020300	0.019025	0.020397
2070	0.010400	0.018325	0.016971	0.016040
2300	0.007928	0.016778	0.015504	0.012317
2530	0.006426	0.015940	0.014724	0.009964
2760	0.005118	0.015024	0.013841	0.007563
2990	0.004420	0.014553	0.013376	0.006033
3220	0.003565	0.013771	0.012660	0.004441
3450	0.003055	0.013507	0.012387	0.003214
3680	0.002865	0.013353	0.012273	0.002722
3910	0.002590	0.013069	0.012011	0.002047
4140	0.002535	0.013076	0.012027	0.001792
4370	0.002556	0.013201	0.012148	0.001809
VIÉS				
230	-0.073529	-0.066282	-0.067147	-0.051349
460	0.058704	0.077895	0.075497	0.093413
690	0.087195	0.114784	0.110881	0.125868
920	0.077563	0.117074	0.112545	0.115869
1150	0.065299	0.114436	0.109553	0.101428
1380	0.055066	0.114590	0.109605	0.086668
1610	0.048204	0.114695	0.109649	0.073828
1840	0.038996	0.111226	0.106329	0.058818
2070	0.032139	0.109157	0.104354	0.046610
2300	0.026585	0.108307	0.103779	0.037237
2530	0.024232	0.108474	0.104146	0.031908
2760	0.022782	0.108350	0.104138	0.027309
2990	0.021241	0.107816	0.103658	0.023535
3220	0.019345	0.106785	0.102780	0.019931
3450	0.018185	0.106554	0.102562	0.017039
3680	0.018218	0.106676	0.102756	0.016356
3910	0.017432	0.106123	0.102273	0.014617
4140	0.017672	0.106394	0.102569	0.014379
4370	0.018258	0.106980	0.103155	0.014965

Figura 7.10 Comportamento dos estimadores km , ek , mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0.0015); Weib(\alpha=0.0015, \beta=1.3); \pi=0.20]$.

Tabela 7.11 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;exp($\lambda=0.0015$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.014296	0.014093	0.013978	0.014990
460	0.032048	0.032600	0.032144	0.034062
690	0.040942	0.042441	0.041519	0.044075
920	0.042076	0.044533	0.043169	0.045855
1150	0.037161	0.040336	0.038565	0.041049
1380	0.029776	0.033566	0.031626	0.033603
1610	0.021210	0.025614	0.023751	0.024465
1840	0.016271	0.020678	0.018959	0.019108
2070	0.011336	0.015680	0.014156	0.013713
2300	0.008240	0.012542	0.011135	0.010210
2530	0.006832	0.011173	0.009842	0.008534
2760	0.004922	0.009304	0.008039	0.006017
2990	0.004085	0.008380	0.007167	0.004853
3220	0.003299	0.007606	0.006441	0.003693
3450	0.002576	0.007066	0.005934	0.002796
3680	0.002181	0.006617	0.005522	0.002155
3910	0.002012	0.006380	0.005309	0.001801
4140	0.001886	0.006288	0.005217	0.001598
4370	0.001621	0.006031	0.004983	0.001164
VIÉS				
230	0.108141	0.107735	0.107167	0.112154
460	0.167406	0.169589	0.168180	0.174241
690	0.190565	0.195379	0.192943	0.199648
920	0.191815	0.199299	0.195706	0.202609
1150	0.177093	0.187682	0.182691	0.189178
1380	0.152728	0.167858	0.161613	0.166020
1610	0.120458	0.142406	0.135561	0.134453
1840	0.100240	0.125804	0.118939	0.113685
2070	0.077679	0.108482	0.101800	0.089637
2300	0.060751	0.096284	0.089699	0.071108
2530	0.050571	0.089930	0.083713	0.059086
2760	0.038906	0.082360	0.076273	0.044574
2990	0.032515	0.078237	0.072369	0.036469
3220	0.026388	0.074234	0.068607	0.028426
3450	0.021900	0.071839	0.066328	0.023061
3680	0.019277	0.069830	0.064478	0.019207
3910	0.017666	0.068460	0.063228	0.016652
4140	0.017010	0.068089	0.062879	0.015625
4370	0.015278	0.066898	0.061808	0.013128

Figura 7.11 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \exp(\lambda=0.0015); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0.40]$.

Tabela 7.12 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;exp($\lambda=0.0015$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
230	0.089754	0.087762	0.087479	0.089344
460	0.045779	0.043937	0.042227	0.047371
690	0.056440	0.058108	0.054514	0.059524
920	0.059928	0.064082	0.058707	0.064230
1150	0.055906	0.065350	0.058568	0.060115
1380	0.052031	0.065273	0.057549	0.055553
1610	0.043927	0.060738	0.052532	0.045409
1840	0.037291	0.058899	0.050611	0.037791
2070	0.032928	0.057858	0.049453	0.031636
2300	0.026177	0.054706	0.046522	0.022968
2530	0.023687	0.054638	0.046267	0.019588
2760	0.022946	0.054713	0.046325	0.018180
2990	0.021301	0.053801	0.045590	0.015134
3220	0.020235	0.053542	0.045340	0.013395
3450	0.020058	0.053733	0.045500	0.012823
3680	0.018292	0.051985	0.043817	0.010174
3910	0.018270	0.052100	0.043938	0.009919
4140	0.017643	0.051791	0.043648	0.008884
4370	0.017725	0.052038	0.043875	0.008942
VIÉS				
230	-0.042598	-0.041017	-0.042801	-0.027998
460	0.152494	0.158967	0.152588	0.169580
690	0.177699	0.202262	0.189840	0.196225
920	0.166029	0.211596	0.196148	0.183517
1150	0.154680	0.218247	0.200743	0.165701
1380	0.140022	0.219704	0.201378	0.144730
1610	0.122371	0.216893	0.197768	0.121172
1840	0.107929	0.215837	0.196933	0.103802
2070	0.097278	0.215053	0.196534	0.088711
2300	0.085750	0.211875	0.194008	0.073036
2530	0.082253	0.213121	0.195345	0.068055
2760	0.080960	0.213615	0.196073	0.065870
2990	0.077588	0.212788	0.195605	0.060462
3220	0.075529	0.212686	0.195648	0.057493
3450	0.075824	0.213323	0.196359	0.057274
3680	0.070929	0.210197	0.193445	0.051120
3910	0.071219	0.210527	0.193827	0.051053
4140	0.070021	0.210143	0.193533	0.049298
4370	0.070607	0.210729	0.194119	0.049884

Figura 7.12 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \exp(\lambda=0.0015); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0.40]$.

Tabela 7.13 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;Weib($\alpha=0.0015,\beta=1.3$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0$].

ERRO QUADRATICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
150	0.017231	0.017507	0.017507	0.017721
300	0.025467	0.026268	0.026268	0.027012
450	0.023385	0.024538	0.024538	0.025646
600	0.017493	0.018743	0.018743	0.019890
750	0.011051	0.012148	0.012148	0.013079
900	0.006613	0.007502	0.007502	0.008188
1050	0.003939	0.004591	0.004591	0.005058
1200	0.002398	0.002826	0.002826	0.003109
1350	0.001473	0.001736	0.001736	0.001933
1500	0.001098	0.001239	0.001239	0.001406
1650	0.000738	0.000806	0.000806	0.000963
1800	0.000533	0.000558	0.000558	0.000714
1950	0.000374	0.000390	0.000390	0.000527
2100	0.000252	0.000261	0.000261	0.000380
2250	0.000175	0.000184	0.000184	0.000286
2400	0.000119	0.000138	0.000138	0.000207
2550	0.000077	0.000110	0.000110	0.000143
2700	0.000047	0.000089	0.000089	0.000092
2850	0.000026	0.000083	0.000083	0.000052
VIÉS				
t	(km)	(ek)	(mk)	(sp)
150	0.116485	0.117828	0.117828	0.119168
300	0.147494	0.150430	0.150430	0.153519
450	0.137499	0.141942	0.141942	0.146536
600	0.113993	0.119724	0.119724	0.125224
750	0.082944	0.089778	0.089778	0.095656
900	0.058293	0.065974	0.065974	0.071471
1050	0.038023	0.046368	0.046368	0.051166
1200	0.022072	0.030946	0.030946	0.034405
1350	0.011382	0.020673	0.020673	0.022666
1500	0.004175	0.013818	0.013818	0.014105
1650	-0.000135	0.009821	0.009821	0.008235
1800	-0.002697	0.007533	0.007533	0.004188
1950	-0.003515	0.006929	0.006929	0.002013
2100	-0.004216	0.006422	0.006422	0.000006
2250	-0.004472	0.006323	0.006323	-0.001253
2400	-0.004239	0.006668	0.006668	-0.002007
2550	-0.003766	0.007216	0.007216	-0.002222
2700	-0.003449	0.007594	0.007594	-0.002433
2850	-0.002938	0.008139	0.008139	-0.002366

Figura 7.13 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \text{Weib}(\alpha=0.0015, \beta=1.3); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0]$.

Tabela 7.14 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;Weib($\alpha=0.0015,\beta=1.3$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=01$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
150	0.210859	0.211132	0.211132	0.209976
300	0.069712	0.071523	0.071523	0.070969
450	0.042191	0.045734	0.045734	0.045879
600	0.036129	0.040321	0.040321	0.041007
750	0.028036	0.032687	0.032687	0.033972
900	0.020642	0.024487	0.024487	0.026355
1050	0.014458	0.017524	0.017524	0.019893
1200	0.011610	0.014285	0.014285	0.016833
1350	0.007216	0.009158	0.009158	0.011588
1500	0.004862	0.006921	0.006921	0.008799
1650	0.003280	0.005140	0.005140	0.006844
1800	0.002404	0.004435	0.004435	0.005157
1950	0.001526	0.003399	0.003399	0.003489
2100	0.001197	0.003340	0.003340	0.002854
2250	0.000811	0.003028	0.003028	0.001928
2400	0.000615	0.002980	0.002980	0.001505
2550	0.000450	0.002914	0.002914	0.001118
2700	0.000260	0.002781	0.002781	0.000736
2850	0.000126	0.002671	0.002671	0.000334
VIÉS				
150	-0.231395	-0.225533	-0.225533	-0.223478
300	0.012494	0.026404	0.026404	0.034022
450	0.106779	0.128783	0.128783	0.138089
600	0.101433	0.130169	0.130169	0.138499
750	0.083744	0.117536	0.117536	0.122118
900	0.055293	0.093778	0.093778	0.092865
1050	0.033343	0.075609	0.075609	0.067874
1200	0.022272	0.067284	0.067284	0.052118
1350	0.006502	0.054129	0.054129	0.030600
1500	0.003335	0.052406	0.052406	0.024030
1650	-0.002055	0.048409	0.048409	0.013877
1800	-0.002297	0.048953	0.048953	0.010387
1950	-0.005915	0.046214	0.046214	0.002868
2100	-0.004056	0.048378	0.048378	0.003003
2250	-0.004232	0.048556	0.048556	0.000454
2400	-0.003279	0.049689	0.049689	0.000329
2550	-0.002726	0.050388	0.050388	-0.000069
2700	-0.002569	0.050671	0.050671	-0.000651
2850	-0.002658	0.050696	0.050696	-0.001747

Figura 7.14 Comportamento dos estimadores km , ek , mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \text{Weib}(\alpha=0.0015, \beta=1.3); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0]$.

Tabela 7.15 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=50;Weib($\alpha=0.0015,\beta=1.3$);Weib($\alpha=0.0015,\beta=1.3$); $\pi=0.40$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
150	0.035101	0.034349	0.034249	0.035218
300	0.066654	0.066575	0.066155	0.067480
450	0.085021	0.085852	0.085042	0.086854
600	0.085916	0.087656	0.086481	0.088649
750	0.077484	0.080016	0.078455	0.080836
900	0.064675	0.068293	0.066488	0.068245
1050	0.051252	0.055342	0.053414	0.054668
1200	0.039251	0.043553	0.041487	0.042224
1350	0.029375	0.033835	0.031673	0.031846
1500	0.021102	0.025377	0.023163	0.023007
1650	0.015988	0.020277	0.018085	0.017446
1800	0.011698	0.016222	0.014115	0.012663
1950	0.009782	0.014436	0.012265	0.010320
2100	0.008234	0.013224	0.011077	0.008316
2250	0.006523	0.011814	0.009732	0.006160
2400	0.005711	0.011237	0.009171	0.004955
2550	0.005008	0.010769	0.008722	0.003931
2700	0.004719	0.010695	0.008637	0.003364
2850	0.004389	0.010637	0.008585	0.002939
VIÉS				
150	0.164766	0.162823	0.162522	0.165365
300	0.252300	0.252454	0.251604	0.254386
450	0.284906	0.286707	0.285271	0.288655
600	0.284936	0.288300	0.286211	0.290429
750	0.269853	0.274998	0.272081	0.276804
900	0.243627	0.251763	0.248043	0.251936
1050	0.213529	0.224182	0.219713	0.222711
1200	0.182869	0.196099	0.190647	0.192324
1350	0.154299	0.170598	0.164118	0.163681
1500	0.125368	0.145544	0.137932	0.133806
1650	0.102189	0.128406	0.120095	0.109415
1800	0.080842	0.113981	0.105158	0.086133
1950	0.068618	0.106852	0.097527	0.071582
2100	0.059983	0.102080	0.092703	0.060701
2250	0.050235	0.096850	0.087416	0.048701
2400	0.045189	0.094409	0.085016	0.041853
2550	0.040322	0.092267	0.082956	0.035389
2700	0.038584	0.091969	0.082677	0.032503
2850	0.036754	0.091675	0.082462	0.030132

Figura 7.15 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=50; \text{Weib}(\alpha=0.0015, \beta=1.3); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0.40]$.

Tabela 7.16 Desempenho dos estimadores com os dados simulados na configuração [N=500;n=10;Weib($\alpha=0.0015, \beta=1.3$);Weib($\alpha=0.0015, \beta=1.3$); $\pi=0.40$].

ERRO QUADRÁTICO MÉDIO				
t	(km)	(ek)	(mk)	(sp)
150	0.216453	0.212942	0.212778	0.215432
300	0.090478	0.088524	0.087317	0.090446
450	0.092595	0.094368	0.091551	0.094018
600	0.098227	0.102763	0.098257	0.099553
750	0.093302	0.099592	0.093606	0.094932
900	0.077946	0.088453	0.081106	0.079790
1050	0.069288	0.083237	0.074905	0.071092
1200	0.056200	0.073296	0.065003	0.057512
1350	0.043145	0.063919	0.055794	0.044373
1500	0.035971	0.061275	0.052691	0.035483
1650	0.033665	0.059885	0.051111	0.031168
1800	0.029382	0.057893	0.049251	0.025132
1950	0.026984	0.056196	0.047890	0.020946
2100	0.025740	0.055744	0.047481	0.017427
2250	0.024060	0.055219	0.046960	0.014983
2400	0.023038	0.055354	0.047083	0.013295
2550	0.023189	0.056256	0.047908	0.012404
2700	0.022739	0.056368	0.048003	0.011381
2850	0.022352	0.056571	0.048159	0.010751
VIÉS				
150	-0.183792	-0.190209	-0.190698	-0.181786
300	0.144727	0.145632	0.142848	0.151299
450	0.256373	0.266677	0.260792	0.266163
600	0.273953	0.291759	0.282603	0.284212
750	0.250429	0.281610	0.269093	0.263210
900	0.216827	0.263856	0.248442	0.228987
1050	0.199592	0.256928	0.239594	0.208099
1200	0.168847	0.243662	0.225954	0.172964
1350	0.138995	0.228933	0.211234	0.140758
1500	0.124317	0.226444	0.208185	0.120689
1650	0.118940	0.224909	0.206573	0.109642
1800	0.106310	0.221265	0.203206	0.093042
1950	0.099885	0.218677	0.201155	0.082698
2100	0.096849	0.218184	0.200902	0.074696
2250	0.093341	0.217627	0.200410	0.069756
2400	0.091737	0.218386	0.201258	0.066949
2550	0.093318	0.220493	0.203363	0.066482
2700	0.092666	0.220856	0.203786	0.064688
2850	0.092386	0.221519	0.204429	0.064038

Figura 7.16 Comportamento dos estimadores km, ek, mk e sp com relação ao viés (EM) e erro quadrático médio (EQM), para a configuração $[N=500; n=10; \text{Weib}(\alpha=0.0015, \beta=1.3); \text{Weib}(\alpha=0.0015, \beta=1.3); \pi=0.40]$.

7.5 Considerações finais.

Conforme mostram as figuras 7.1 a 7.16, a posição de cada estimador, em relação aos demais, quanto ao erro quadrático médio e viés, não se altera nas diferentes configurações simuladas. Entretanto, a magnitude do EQM e EM aumenta sensivelmente, para todos os estimadores, com a diminuição do tamanho da amostra e/ou com o aumento na porcentagem de censuras e/ou com a escolha do tipo de distribuição para os tempos T_e e T_o e/ou com a escolha do valor dos parâmetros dessas distribuições.

De um modo geral, tanto o EQM quanto o EM são maiores no início e depois decaem lentamente com o passar do tempo.

Com as censuras, o estimador modificado apresenta sempre um desempenho ligeiramente melhor que o do estimador de Kitchin. Em algumas configurações ele chega a superar, na primeira metade da curva, o desempenho do estimador de Salinas e Pereira (veja-se, por exemplo as figuras 7.1, 7.3, 7.7, 7.9, 7.11, 7.13 e 7.15).

O estimadores de Kaplan e Meier e de Salinas e Pereira são, nesta ordem, os que mais apresentam viés negativo nos extremos da curva (veja-se, por exemplo, as tabelas 7.1, 7.2, 7.6, 7.12 e 7.13). A quantidade de viés negativo nos estimadores de Kitchin e modificado de Kitchin é muito menor, indicando que as suas curvas estimadas têm a tendência de serem, em média, mais "altas" que as daquelas (como era de se esperar, pelo que já apontavam as desigualdades A) e B) de (6.6).

De um modo geral, os resultados das simulações indicam a classificação dos estimadores estudados, segundo a ordem dos melhores desempenhos, em dois grupos: no primeiro aparecem os estimadores de Kaplan e Meier e o de Salinas e Pereira e no segundo, o estimador Modificado de Kitchin e o estimador de Kitchin.

CONCLUSÕES:

Todos os estimadores F-N-P abordados neste trabalho podem ser expressos em termos de funções empíricas. A interpretação geométrica da taxa instantânea de Kitchin foi estendida aos demais estimadores F-N-P, sendo identificadas duas classes de estimadores. A primeira, em que as taxas acumuladas nos subintervalos foram interpretadas como uma razão de áreas e a segunda, como logaritmos de uma razão de áreas. Essa interpretação geométrica permitiu uma comparação analítica entre as taxas dos estimadores F-N-P, abrindo espaço para a definição do estimador alternativo proposto.

Esse estimador apresenta uma taxa cujo cálculo é mais simplificado do que o da taxa de Kitchin, utilizando somente contagens dos indivíduos em risco, mortos e censurados. Além disso, ele apresentou, na simulação, um desempenho ligeiramente melhor do que o de Kitchin.

Os desempenhos desses estimadores não traduziram as vantagens esperadas devidas à sua continuidade. Aqui, como no trabalho de Kitchin, observou-se que quando as censuras estão presentes, o estimador de Kitchin apresenta significantes vícios positivos na cauda direita da função básica e esse vício aumenta com a intensidade das censuras. Isso também aconteceu com o novo estimador proposto.

Ambos os estimadores apresentam uma tendência a superestimar as probabilidades de sobrevivência nos extremos de cada subintervalo (conforme mostra a comparação analítica das taxas).

O estimador de Kaplan e Meier, além da sua simplicidade de cálculo, foi o que apresentou o melhor desempenho e portanto continua sendo o mais recomendado nos tipos de problemas aqui mencionados. Este estimador também apresenta a vantagem de se encontrar implantado nos principais programas estatísticos atualmente em uso.

Mesmo sob uma priori não informativa, o estimador de Salinas e Pereira mostrou um razoável desempenho a partir da segunda me-

tade da curva básica. Entretanto, além da dificuldade na definição da priori, ele exige um esforço computacional muito maior do que os outros. Outra desvantagem é que suas estimativas também são afetadas pelos métodos numéricos necessários para o seu cálculo.

Para dados que se ajustem às situações aqui testadas, os resultados apontam para os estimadores de Kaplan e Meier, de Salinas e Pereira, modificado de Kitchin e de Kitchin, nessa ordem. Por outro lado, aliando-se simultaneamente o desempenho e a simplicidade operacional, os melhores estimadores são, pela ordem, os de Kaplan e Meier e modificado de Kitchin.

Fora das circunstâncias aqui estudadas, não existe uma "receita" que indique qual estimador seja o mais apropriado. Entretanto o material contido neste trabalho pode servir de base para uma reflexão e facilitar a opção por uma alternativa que se julgue mais aplicável aos dados e condições que se disponha no momento.

Com o emprego do raciocínio aqui desenvolvido pode-se continuar a busca de outros estimadores para a curva de sobrevivência sob dados censurados.

Em síntese, este trabalho reúne e acrescenta conhecimentos julgados úteis para auxiliar o profissional de saúde e de biometria na análise de dados censurados de sobrevivência, desenvolvendo uma metodologia própria de comparação dos principais métodos de estimação.

BIBLIOGRAFIA E ANEXO

REFERÊNCIAS BIBLIOGRÁFICAS

01. AALEN, O. (1976) - Nonparametric inference in connection with multiple decrement models. Scand. J. Stat., 3, 15-27.
02. AALEN, O. (1978a) - Nonparametric inference for a family of counting processes. Ann. Stat., 4, 701-726.
03. AALEN, O. ; JHOANSEN, S. (1978) - An empirical transition matrix for non-homogeneous Markov chains based on censored observations Scand. J. Stat., 5, 141-150.
04. AALEN, O. (1981) - Practical applications of the nonparametric statistical theory for counting processes. Technical Report of Institute of Mathematical and Physical Sciences - University of Tromso, Norway.
05. ABRAMOWITZ, M. e STEGUN, I. A. (1972). Handbook of mathematical functions. Dover Publications, Inc, New York.
06. 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.
07. BARRETO, M. C. M. (1982) - Análise de curvas de sobrevivência- O modelo de Cox. Dissertação de mestrado, IME-USP.
08. BAYES, T. (1953) - An essay towards solving a problem in the doctrine of chances. Biometrika, 45, 293-315 (reprinted from Philos. Trans. Roy. Soc. London, 53, 1763, 370-418, with biographical note by G. A. Barnard.
09. BEKMAN, O. R. e COSTA NETO, P. L. O. (1980). Análise estatística da decisão. Editora Edgard Blücher Ltda, São Paulo, SP, Brasil.
10. BENZE, B. G. (1985) - Estimadores análogos ao de Kaplan-Meier em modelos de riscos competitivos generalizados. Dissertação de mestrado, Faculdade de Medicina-Ribeirão Preto-USP.
11. BENZE, B. G. e FRANCO, M. A. P. (1985) - Estimadores análogos ao de Kaplan-Meier em modelos de riscos competitivos generalizados. Anais da 5a. Reunião Regional da Sociedade Brasileira de Ma-

temática Aplicada e Computacional. UNESP, Rio Claro, 186-191.

12. BERLINER, M. ; HILL, B. (1988) - Bayesian nonparametric survival analysis. JASA , 83 , 772-779.
13. BERQUÓ, E. S. , SOUZA, J. M. P. de, GOTLIEB, S. L. D. (1981). Bioestatística. Editora Pedagógica e Universitária Ltda, São Paulo.
14. BIRBAUM, Z. W. (1979) - On the mathematics of competing risks. Vital and Health Statistics. Series 2, No.77, 60p.
15. BRASIL, G. H. (1991) - Estatística Bayesiana: alguns aspectos teóricos e práticos. XXII Reunião Regional de Estatística, UFPE, Recife, PE.
16. BRESLOW, N. ; CROWLEY, J. (1974) - A large sample study of the life table and product limit estimates under random censorship. Ann. Stat. , 2 , 437-453.
17. BRUCE, W. ; TURNBULL JR, B. ; BROWN, W. ; HU, M. (1974) - Survivorship of heart transplant data. J. Am. Stat. Ass. , 69 , 74-80.
18. CANTOR, A. B. ; SHUSTER, J. J. (1992) - Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Stat Med , 11 , 931-937.
19. CHIANG, C. L. (1960a) - A stochastic study of the life table and its applications : I. Probability distributions of the biometric functions. Biometrics , 16 , 618-635.
20. CHIANG, C. L. (1968) - Introduction to stochastic processes in biostatistics. John Wiley, New York.
21. CNAAN, A. ; RYAN, L. (1989) - Survival analysis in natural history studies of disease. Stat Med , 8 , 1255-1268.
22. CORNFIELD, J. (1957) - The estimation of the probability of developing a disease in the presence of competing risks. Amer. J. Publ. Hlth. , 47 , 601-607.
23. CUTLER, S. J. ; EDERER, F. (1958) - Maximum utilization of the life table method in analyzing survival. Journal of Chronic Diseases 8, 699-712.
24. ESTEVE, J. ; BENHAMOU, E. ; CROASDALE, M. ; RAYMOND, L. (1990) - Relative survival and the estimation of net survival: elements for fur

ther discussion. Stat Med , 9 , 529-538.

25. FLEMING, T. R. (1978a) - Nonparametric estimation for non-homogeneous Markov processes in the problem of competing risks. Ann. Stat. , 6 , 1057-1070.
26. FLEMING, T. R. (1978b) - Asymptotic distribution results in competing risks estimation. Ann. Stat. , 6 , 1071-1079.
27. GAIL, M. (1975) - A review and critique of some models used in competing risk analysis. Biometrics , 31 , 209-222.
28. GILL, R. D. ; JOHANSEN, S. (1990) - A survey of product-integration with a view toward application in survival analysis. Ann. Stat. , 18 , 1501-1555.
29. GOTLIEB, S. L. D. (1977) - Mortalidade diferencial por causas São Paulo, 1970- Tábua de vida de múltiplo decremento. Tese de doutorado , Fac. Saúde Pública-USP.
30. HILLIS, A. ; MAGUIRRE, M. ; HAWKINS, B. S. e NEWHOUSE, M. M. (1986) - The Markov process as a general method for nonparametric analysis of right-censored medical data. J. Chron. Dis. , 39 , 595-604.
31. HJORT, N. L. (1990) - Nonparametric Bayes estimators based on beta processes in models for life history data. Ann. Stat. , 18 , 1259-1294.
32. HOEL, D. G. ; WALBURG, H. E. (1972) - Statistical analysis of survival experiments. J. Natl. Cancer Inst. , 49 , 361-372.
33. KAPLAN, E. L. ; MEIER, P. (1958) - Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. , 58 , 457-481.
34. KITCHIN, J. (1980) - A new method for estimating life distributions from incomplete data. Doctoral Dissertation , Department of Statistics, Florida State University.
35. KLEIN, J. P. ; LEE, S. -C. , MOESCHBERGER, M. L. (1990) - A partially parametric estimator of survival in the presence of randomly censored data. Biometrics , 46 , 795-811.
36. KLEINBAUM, D. G. ; KUPPER L. L. and MORGENSTERN H. (1982) - Epidemiologic research: principles and quantitative methods. Lifetime Learning Publ. , Belmont, Calif. 519p.

37. LEE, E. T. (1980) - Statistical methods for survival data analysis. Lifetime Learning Publications, Belmont, California.
38. LIMA, A. C. P. de (1988). Estimação em modelos de sobrevivência. Dissertação de Mestrado - IME-USP.
39. LINDLEY, D. V. (1983) - Lectures on Bayesian statistics. Publicações do IME-USP.
40. LINK, W. A. (1989) - A model for informative censoring. JASA , 84, 749-552.
41. MARIOTTO, A. B. ; MARIOTTI, S. ; PEZZOTTI, P. ; REZZA, G. ; VERDECHIA, A. (1992) - Estimation of the acquired immunodeficiency syndrome incubation period in intravenous drug users : a comparison with homosexuals. Am. J. Epid. , 135, 428-437.
42. MILLER, JR., R. G. ; GONG, G. ; MUÑOZ, A. (1981). Survival Analysis. John Wiley & Sons Inc.
43. MILLER, JR., R. G. (1983) - What price Kaplan-Meier? Biometrics , 39 1077-1081.
44. MOESCHBERGER, M. L. ; DAVID, H. A. (1971) - Life tests under competing causes of failure and the theory of competing risks. Biometrics , 27 , 909-933.
45. MOESCHBERGER, M. L. ; KLEIN, J. P. (1985) - A comparison of several methods of estimating the survival function when there is extreme right censoring. Biometrics , 41 , 253-259.
46. NELSON, W. (1969) - Hazard plotting for incomplete failure data. J. Qual. Tech. , 1 , 27-52.
47. NELSON, W. (1970) - Hazard plotting methods for analysis of life data with different failure modes. J. Qual. Tech. , 2 , 126-149.
48. OLIVEIRA, L. A. DE (1981). Teste do tipo Wilcoxon para comparação de curvas de sobrevivência. Dissertação de Mestrado - Instituto de Ciências Matemáticas de São Carlos, USP.
49. OLIVEIRA, L. A. e BARBOSA, E. P. (1983) - Uma comparação entre os estimadores produto-limite (Kaplan-Meier) e exponencial por partes (Kitchin). Congresso da Sociedade Brasileira de Biometria - Jaboticabal-SP.

50. PAES, N. A. (1982) - Tábuas de vida de múltiplo decremento modelo de Chiang. Dissertação de mestrado, IME-USP.
51. PEPE, M. S. ; MORI, M. (1993) - Kaplan-Meier marginal or conditional probability curves in summarizing competing risks failure time data. Stat Med , 12 , 737-751.
52. PEREIRA, C. A. R. e VIANA, M. A. G. (1982) - Elementos de inferência Bayesiana. 5^o Simpósio Nacional de Probabilidade e Estatística.
53. PETERSON JR, A. V. (1975) - Nonparametric estimation in the competing risks problem. Tech. report no 73, Stanford Univ.
54. PETERSON JR, A. V. (1976) - Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci, USA, 73 , 11-13.
55. PETERSON JR, A. V. (1977) - Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. J. Am. Stat. Ass. 360 , 854-858.
56. PRENTICE, R. L. (1975) - Models for survival with censoring and explanatory variables. Biometrika , 60 , 279-288.
57. RAGLAND, K. E. ; SELVIN, S. ; MERRILL, D. W. (1991) - Black-white differences in stage-specific cancer survival: analysis of seven selected sites. Am. J. Epid. , 133 , 672-682.
58. ROBERTSON, J. B. ; UPPULURI, V. R. R. (1984) - A generalized Kaplan-Meier estimator. Ann. Stat. , 12 , 366-371.
59. SANTOS, J. L. (1972) - Contribuição para o tratamento estocástico da tabela de sobrevivência e suas aplicações. Tese de doutorado, Fac. Saúde Pública-USP.
60. SANTOS, J. L. F. (1972) - Sobre a força de mortalidade na construção de tábuas de sobrevivência. Rev Saúde Publ , 6 : 263-67.
61. SOARES, J. F. e BARTMAN, F. C. (1983) - Métodos estatísticos em medicina e biologia. 14o. Colóquio Brasileiro de Matemática.
62. SOUZA, J. M. P. de (1984) - Análise de metodologia em estudos epidemiológicos de câncer - comparações de técnicas probabilísticas. Tese de livre-docência, Fac. Saúde Pública-USP.

63. STRUTHERS, C. A. ; FAREWELL, V. T. (1989) - A mixture model for times to AIDS data with left truncation and an uncertain origin. Biometrika , 76 , 814-817.
64. SUZARLA, V. ; VAN RYZIN, J. (1976) - Nonparametric Bayesian estimation of survival curves from incomplete observation. J. Amer. Stat. Assoc. 71 , 897-902.
65. SUZARLA, V. ; VAN RYZIN, J. (1978b) - Large Sample Theory for a Bayesian Non-parametric Survival Curve Estimator Based on Censored Samples. Ann. Statist. 6, 755-768.
Addendum (1980), Ann. Statist. 8, 683.
66. THOMAS, D. R. , (1972). On a generalized savage statistic for comparing two arbitrarily censored samples. Technical Report, Oregon State Univ.
67. TISIATIS, A. (1975) - A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. USA, 72, 20-22.
68. TSAI, W. -Y. (1986) - Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. Ann. Stat. 14 , 238-249.
69. TU, X. M. ; MENG, X. -L. ; PAGANO, M. (1993) - The AIDS epidemic: estimating survival after AIDS diagnosis from surveillance data. JASA , 88 , 26-35.
70. SALINAS-TORRES, V. H. (1992) - Análise Bayesiana não-paramétrica na teoria de riscos competitivos. Tese de Doutorado-IME-USP.
71. SALINAS-TORRES, V. H. ; PEREIRA, C. A. e TIWARI, R. C. (1992) - A non-parametric Bayesian analysis of competing risks models. Relatório Técnico, IME-USP.

ANEXO

Tabela A1 Entradas e saídas, durante a observação de uma amostra de pacientes de transplante de rim, na Unidade de Transplantes Renais do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, entre 14/02/72 a 31/12/78 (doador vivo e esquema terapêutico A) e duração dos intervalos de seguimento.

ORDEM DE EN- TRADA	DO TRANS- PLANTE	DATA		No. DE DIAS		No. DE		INTERVALO DE SEGUIMENTO EM DIAS
		DA ULTI- MA VERI- FICAÇÃO	DO INÍCIO DO SEGUIMENTO A- TÉ A CIRURGIA	DA CIRURGIA ATÉ O TRUN- CAMENTO	MESES	no. dias		
						30,44		
01	12/04/72	13/05/73	58 **	396	13,01			[58; 454]
02	24/05/72	18/10/77	100	1973	64,82			[100; 2073]
03	31/05/72	01/01/73	107	215	7,06			[107; 322]
04	28/06/72	15/10/76	135	1570	51,58			[135; 1705]
05	26/07/72	12/02/73	163	201	6,60			[163; 364]
06	23/04/74	31/12/78	799	1713 *	56,27			[799; 2512]
07	05/08/74	31/12/78	903	1609 *	52,86			[903; 2512]
08	26/08/74	31/12/78	924	1588 *	52,17			[924; 2512]
09	17/09/74	15/12/74	946	89	2,92			[946; 1035]
10	23/09/74	02/11/74	952	40	1,31			[952; 992]
11	14/10/74	07/09/78	973	1424	46,78			[973; 2397]
12	04/11/74	31/12/78	994	1518 *	49,87			[994; 2512]
13	09/12/74	31/12/78	1029	1483 *	48,72			[1029; 2512]
14	13/01/75	08/06/76	1064	512	16,82			[1064; 1576]
15	27/01/75	31/12/78	1078	1434 *	47,11			[1078; 2512]
16	04/02/75	31/12/78	1086	1426 *	46,85			[1086; 2512]
17	09/04/75	31/12/78	1150	1362 *	44,74			[1150; 2512]
18	28/04/75	31/12/78	1169	1343 *	44,12			[1169; 2512]
19	12/05/75	20/09/78	1183	1227	40,31			[1183; 2410]
20	08/07/75	31/12/78	1240	1272 *	41,79			[1240; 2512]
21	07/04/76	01/10/77	1514	542	17,80			[1514; 2056]
22	28/04/76	10/06/76	1535	43	1,41			[1535; 1578]
23	05/05/76	31/12/78	1542	970 *	31,87			[1542; 2512]
24	19/05/76	31/12/78	1556	956 *	31,41			[1556; 2512]
25	14/07/76	18/05/77	1612	308	10,12			[1612; 1920]

* refere-se às observações censuradas e o valor 30,44 é especificado na sexta coluna, levando-se em consideração os anos bissextos.
 ** os valores foram obtidos pelo procedimento descrito a seguir.

Para a obtenção dos valores da quarta coluna da tabela A1, contabilizou-se os meses do ano, com as suas quantidades de dias:

01	JANEIRO	31 dias
02	FEVEREIRO	28 (ou 29 dias para os anos bissextos)
03	MARÇO	31 dias
04	ABRIL	30 dias
05	MAIO	31 dias
06	JUNHO	30 dias
07	JULHO	31 dias
08	AGOSTO	31 dias
09	SETEMBRO	30 dias
10	OUTUBRO	31 dias
11	NOVEMBRO	30 dias
12	DEZEMBRO	31 dias

e construiu-se a seguinte tabela:

Tabela A2 Contagem do tempo (em dias) de seguimento dos pacientes de transplantes renais da tabela A1.

ORDEM DE EN- TRADA	A N OS	CONTAGEM DE DIAS A PARTIR DE 14/02/72 ATÉ O DIA ANTERIOR À DATA EXPOSTA NA SEGUNDA COLUMNA DA TAB. 1					TOTAL
		1972	1973	1974	1975	1976	
01		16+31+11					58
02		16+61+23					100
03		16+61+30					107
04		16+92+27					135
05		16+122+25					163
06		16+306	365	90+22			799
07		16+306	365	212+04			903
08		16+306	365	212+25			924
09		16+306	365	243+16			946
10		16+306	365	243+22			952
11		16+306	365	273+13			973
12		16+306	365	304+03			994
13		16+306	365	334+08			1029
14		16+306	365	365	12		1064
15		16+306	365	365	26		1078
16		16+306	365	365	31+03		1086
17		16+306	365	365	90+08		1150
18		16+306	365	365	90+27		1169
19		16+306	365	365	120+11		1183
20		16+306	365	365	181+07		1240
21		16+306	365	365	365	91+06	1514
22		16+306	365	365	365	91+27	1535
23		16+306	365	365	365	121+04	1542
24		16+306	365	365	365	121+18	1556
25		16+306	365	365	365	182+13	1612

Tabela A3 Cálculo das estimativas das curvas de subsobrevivência e de sobrevivência "bruta" e da "média bruta", para os dados do exemplo de transplante de rim, doador vivo, esquema A, na UTR do Hospital das Clínicas de Ribeirão Preto.

TEMPOS DE DURAÇÃO	n S1 _T (t)	n S2 _T (t)	n S _T (t)	t _m - t _{m-1} = Δ _m	n S _T (t). Δ _m = área _m
0	0,52	0,48	1,00	40 - 0 = 40	40,00
40	0,48	0,48	0,96	43 - 40 = 3	2,88
43	0,44	0,48	0,92	89 - 43 = 46	43,32
89	0,40	0,48	0,88	201 - 89 = 112	98,56
201	0,36	0,48	0,84	215 - 201 = 14	11,76
215	0,32	0,48	0,80	308 - 215 = 93	74,40
308	0,28	0,48	0,76	396 - 308 = 88	66,88
396	0,24	0,48	0,72	512 - 396 = 116	83,52
512	0,20	0,48	0,68	542 - 512 = 30	20,40
542	0,16	0,48	0,64	956 - 542 = 414	264,96
956 *	0,16	0,44	0,60	970 - 956 = 14	8,40
970 *	0,16	0,40	0,56	1227 - 970 = 257	143,92
1227	0,12	0,40	0,52	1272 - 1227 = 45	23,40
1272 *	0,12	0,36	0,48	1343 - 1272 = 71	34,08
1343 *	0,12	0,32	0,44	1362 - 1343 = 19	8,36
1362 *	0,12	0,28	0,40	1424 - 1362 = 62	24,80
1424	0,08	0,28	0,36	1426 - 1424 = 2	0,72
1426 *	0,08	0,24	0,32	1434 - 1426 = 8	2,56
1434 *	0,08	0,20	0,28	1483 - 1434 = 49	13,72
1483 *	0,08	0,16	0,24	1518 - 1483 = 35	8,40
1518 *	0,08	0,12	0,20	1570 - 1518 = 52	10,40
1570	0,04	0,12	0,16	1588 - 1570 = 18	2,88
1588 *	0,04	0,08	0,12	1609 - 1588 = 21	2,52
1609 *	0,04	0,04	0,08	1713 - 1609 = 104	8,32
1713 *	0,04	0,00	0,04	1973 - 1713 = 260	10,40
1973	0,00	0,00	0,00	<div> <div> <div>25</div> <div>Σ (áreas)_m = 1008,56</div> </div> </div>	

Tabela A4 Estimativa atuarial (eal), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

x-Δ	x	l_x	d_x	u_x	w_x	l'_x	q_x	p_x	ea S_x	área em meses x
0	- 1	25	-	-	-	25,0	-	1	1,0000	1,0000
1	- 2	25	2	-	-	25,0	0,0800	0,9200	1,0000	1,0000
2	- 3	23	1	-	-	23,0	0,0435	0,9565	0,9200	0,9200
3	- 4	22	-	-	-	22,0	0	1	0,8800	
4	- 5	22	-	-	-	22,0	0	1	0,8800	3,5200
5	- 6	22	-	-	-	22,0	0	1	0,8800	
6	- 7	22	1	-	-	22,0	0,0455	0,9545	0,8800	
7	- 8	21	1	-	-	21,0	0,0476	0,9524	0,8400	0,8400
8	- 9	20	-	-	-	20,0	0	1	0,8000	
9	-10	20	-	-	-	20,0	0	1	0,8000	2,4000
10	-11	20	1	-	-	20,0	0,0500	0,9500	0,8000	
11	-12	19	-	-	-	19,0	0	1	0,7600	
12	-13	19	-	-	-	19,0	0	1	0,7600	2,2800
13	-14	19	1	-	-	19,0	0,0526	0,9474	0,7600	
14	-15	18	-	-	-	18,0	0	1	0,7200	
15	-16	18	-	-	-	18,0	0	1	0,7200	2,1600
16	-17	18	1	-	-	18,0	0,0556	0,9444	0,7200	
17	-18	17	1	-	-	17,0	0,0588	0,9412	0,6800	0,6800
18	-19	16	-	-	-	16,0	0	1	0,6400	
...	...	...	...	...	...	...	...	...	...	
30	-31	16	-	-	-	16,0	0	1	0,6400	
31	-32	16	-	-	2	15,0	0	1	0,6400	
32	-33	14	-	-	-	14,0	0	1	0,6400	14,7200
...	...	...	...	...	...	...	...	...	...	
39	-40	14	-	-	-	14,0	0	1	0,6400	
40	-41	14	1	-	-	14,0	0,0714	0,9286	0,6400	
41	-42	13	-	-	1	12,5	0	1	0,5943	
42	-43	12	-	-	-	12,0	0	1	0,5943	
43	-44	12	-	-	-	12,0	0	1	0,5943	3,5658
44	-45	12	-	-	2	11,0	0	1	0,5943	
45	-46	10	-	-	-	10,0	0	1	0,5943	
46	-47	10	1	-	1	9,5	0,1053	0,8947	0,5943	
47	-48	8	-	-	1	7,5	0	1	0,5317	
48	-49	7	-	-	1	6,5	0	1	0,5317	
49	-50	6	-	-	1	5,5	0	1	0,5317	2,6585
50	-51	5	-	-	-	5,0	0	1	0,5317	
51	-52	5	1	-	-	5,0	0,2000	0,8000	0,5317	
52	-53	4	-	-	2	3,0	0	1	0,4254	
53	-54	2	-	-	-	2,0	0	1	0,4254	
54	-55	2	-	-	-	2,0	0	1	0,4254	
55	-56	2	-	-	-	2,0	0	1	0,4254	5,5302
56	-57	2	-	-	1	1,5	0	1	0,4254	
57	-58	1	-	-	-	1,0	0	1	0,4254	
...	...	...	...	...	...	...	...	...	...	
63	-64	1	-	-	-	1,0	0	1	0,4254	
64	-65	1	1	-	-	1,0	1	0	0,4254	
65	-								0,0000	Σ =41,2745
* Σ(em dias) = 1257,2										

Tabela A5 Estimativa atuarial (ea2), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

$t_{m-1} - t_m$	$V[t_{m-1}]$	$M[I_m]$	$W[I_m]/2$	$p1[I_m]$	$S(t)$	Δ_m^1	área_m^1 (em dias)
0 - 40	25	1	0	0,9600	1,0000	40	40,000
40 - 43	24	1	0	0,9583	0,9600	3	2,880
43 - 89	23	1	0	0,9565	0,9200	46	42,320
89 - 201	22	1	0	0,9545	0,8800	112	98,560
201 - 215	21	1	0	0,9524	0,8400	14	11,760
215 - 308	20	1	0	0,9500	0,8000	93	74,400
308 - 396	19	1	0	0,9474	0,7600	88	66,880
396 - 512	18	1	0	0,9444	0,7200	116	83,520
512 - 542	17	1	0	0,9412	0,6800	30	20,400
542 -1227	16	1	2/2	0,9333	0,6400	685	436,688
1227 -1424	13	1	3/2	0,9130	0,5973	197	115,540
1424 -1570	09	1	4/2	0,8571	0,5453	146	74,925
1570 -1973	4	1	3/2	0,6000	0,4674	403	147,725
1973 -					0,2805		
$\Sigma(\text{dias})=1215,600$							

Tabela A6 Estimativa de Kaplan e Meier (km), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

ORDEM j	TEMPO t em dias	$\frac{n-j}{n-j+1}$	km = $p1_j(t)$	$\prod_{t1_j \leq t}^{km} p1_j(t) = S1(t)$	area ¹ _j em dias
	0		=1,0	1,0 =1,0000	
01	40	(25-1)/(25-1+1)	=0,9600	1,0000x0,9600=0,9600	40,0000
02	43	(25-2)/(25-2+1)	=0,9583	0,9600x0,9583=0,9200	2,8800
03	89	(25-3)/(25-3+1)	=0,9565	0,9200x0,9565=0,8800	42,3200
04	201	(25-4)/(25-4+1)	=0,9545	0,8800x0,9545=0,8400	98,5600
05	215	(25-5)/(25-5+1)	=0,9524	0,8400x0,9524=0,8000	11,7600
06	308	(25-6)/(25-6+1)	=0,9500	0,8000x0,9500=0,7600	74,4000
07	396	(25-7)/(25-7+1)	=0,9474	0,7600x0,9474=0,7200	66,8800
08	512	(25-8)/(25-8+1)	=0,9444	0,7200x0,9444=0,6800	83,5200
09	542	(25-9)/(25-9+1)	=0,9412	0,6800x0,9412=0,6400	20,4000
10	956*	1	= 1	0,6400x1,0 =0,6400	
11	970*	1	= 1	0,6400x1,0 =0,6400	
12	1227	(25-12)/(25-12+1)	=0,9286	0,6400x0,9286=0,5943	438,4000
13	1272*	1	= 1	0,5943x1,0 =0,5943	
14	1343*	1	= 1	0,5943x1,0 =0,5943	
15	1362*	1	= 1	0,5943x1,0 =0,5943	
16	1424	(25-16)/(25-16+1)	=0,9000	0,5943x0,9000=0,5349	117,0771
17	1426*	1	= 1	0,5349x1,0 =0,5349	
18	1434*	1	= 1	0,5349x1,0 =0,5349	
19	1483*	1	= 1	0,5349x1,0 =0,5349	
20	1518*	1	= 1	0,5349x1,0 =0,5349	
21	1570	(25-21)/(25-21+1)	=0,8000	0,5349x0,8000=0,4279	78,0954
22	1588*	1	= 1	0,4279x1,0 =0,4279	
23	1609*	1	= 1	0,4279x1,0 =0,4279	
24	1713*	1	= 1	0,4279x1,0 =0,4279	
25	1973	(25-25)/(25-25+1)	= 0	0,4279x 0 =0	172,4437
					Σ=1246,4437

Tabela A7 Estimativa de Nelson (en), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

OR DEM j	TEMPO t em dias	$\frac{1}{n-j+1}$	en $=r(I_m)$	$\sum_{t_1 \leq t} r_1(t) \Delta t_1 = R_1(t)$	en $S_1(t)$	$\overset{1}{\text{área}}_m$ dias
	0		=0	0	=0	1,0000
01	40	$1/(25-1+1)$	=0,0400	$0,0000+0,0400=0,0400$	0,9608	40,0000
02	43	$1/(25-2+1)$	$\cong 0,0417$	$0,0400+0,0435\cong 0,0817$	0,9216	2,8824
03	89	$1/(25-3+1)$	$\cong 0,0435$	$0,1252+0,0455\cong 0,1252$	0,8824	42,3936
04	201	$1/(25-4+1)$	$\cong 0,0455$	$0,1707+0,0476\cong 0,1707$	0,8432	98,8288
05	215	$1/(25-5+1)$	$\cong 0,0476$	$0,2183+0,0500\cong 0,2183$	0,8039	11,8048
06	308	$1/(25-6+1)$	=0,0500	$0,2683+0,0526\cong 0,2683$	0,7647	74,7627
07	396	$1/(25-7+1)$	$\cong 0,0526$	$0,3209+0,0556\cong 0,3209$	0,7255	67,2936
08	512	$1/(25-8+1)$	$\cong 0,0556$	$0,3765+0,0588\cong 0,3765$	0,6863	84,1580
09	542	$1/(25-9+1)$	$\cong 0,0588$	$0,4353+0,0 \cong 0,4353$	0,6471	20,5890
10	956*	0	=0	$0,4353+0,0 \cong 0,4353$	0,6471	
11	970*	0	=0	$0,4353+0,0 \cong 0,4353$	0,6471	
12	1227	$1/(25-12+1)$	$\cong 0,0714$	$0,5067+0,0714\cong 0,5067$	0,6025	443,2635
13	1272*	0	=0	$0,5067+0,0 \cong 0,5067$	0,6025	
14	1343*	0	=0	$0,5067+0,0 \cong 0,5067$	0,6025	
15	1362*	0	=0	$0,5067+0,0 \cong 0,5067$	0,6025	
16	1424	$1/(25-16+1)$	=0,1000	$0,5067+0,1000\cong 0,6067$	0,5452	118,6925
17	1426*	0	=0	$0,6067+0,0 \cong 0,6067$	0,5452	
18	1434*	0	=0	$0,6067+0,0 \cong 0,6067$	0,5452	
19	1483*	0	=0	$0,6067+0,0 \cong 0,6067$	0,5452	
20	1518*	0	=0	$0,6067+0,0 \cong 0,6067$	0,5452	
21	1570	$1/(25-21+1)$	=0,2000	$0,6067+0,2000\cong 0,8067$	0,4463	79,5992
22	1588*	0	=0	$0,8067+0,0 \cong 0,8067$	0,4463	
23	1609*	0	=0	$0,8067+0,0 \cong 0,8067$	0,4463	
24	1713*	0	=0	$0,8067+0,0 \cong 0,8067$	0,4463	
25	1973	$1/(25-25+1)$	=1,0000	$0,8067+1,0 \cong 1,8067$	0,1642	179,8589
						$\Sigma \cong 1264,1270$

Obs: I_m é o subintervalo entre $t_{1_{m-1}}$ e t_{1_m} e $\Delta t_1 = [t_{1_m} - t_{1_{m-1}}]$

Tabela A8 Estimativa de Kitchin (e_k), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

TEMPO	$(t1_m - t1_{m-1})$	e_k $r1_m$	e_k $R1(t) +$ $(x - t1_{m-1})r1$	e_k $S1(t)$	# área ¹ _m em dias
0				1,0000	
40	(0; 40]	$(25(40 - 0))^{-1}$ $\cong 0,0010$	0,0400	0,9608	39,211
43	(40; 43]	$(24(43 - 40))^{-1}$ $\cong 0,0139$	0,0400+0,0417	0,9216	2,823
89	(43; 89]	$(23(89 - 43))^{-1}$ $\cong 0,0009$	0,0817+0,0435	0,8824	41,526
201	(89; 201]	$(22(201 - 89))^{-1}$ $\cong 0,0004$	0,1251+0,0454	0,8432	96,649
215	(201; 215]	$(21(215 - 201))^{-1}$ $\cong 0,0035$	0,1706+0,0476	0,8039	11,520
308	(215; 308]	$(20(308 - 215))^{-1}$ $\cong 0,0005$	0,2182+0,0500	0,7647	73,057
396	(308; 396]	$(19(396 - 308))^{-1}$ $\cong 0,0006$	0,2682+0,0526	0,7255	65,553
512	(396; 512]	$(18(512 - 396))^{-1}$ $\cong 0,0005$	0,3208+0,0556	0,6863	81,772
542	(512; 542]	$(17(542 - 512))^{-1}$ $\cong 0,0020$	0,3764+0,0588	0,6471	19,984
956*		$(16(956 - 542) +$	0,4352+0,0397	0,6219	
970*	(542; 1227]	$15(970 - 956) +$	0,4352+0,0410	0,6211	
1227		$14(1227 - 970))^{-1}$ $\cong 0,0001$	0,4352+0,0657	0,6060	428,446
1272*		$(13(1272 - 1227) +$	0,5009+0,0199	0,5941	
1343*		$12(1343 - 1272) +$	0,5009+0,0512	0,5757	
1362*	(1227; 1424]	$11(1362 - 1343) +$	0,5009+0,0596	0,5709	
1424		$10(1424 - 1362))^{-1}$ $\cong 0,0004$	0,5009+0,0869	0,5555	114,797
1426*		$(9(1426 - 1424) +$	0,5878+0,0022	0,5543	
1434*		$8(1434 - 1426) +$	0,5878+0,0112	0,5494	
1483*	(1424; 1570]	$7(1483 - 1434) +$	0,5878+0,0659	0,5201	
1518*		$6(1518 - 1483) +$	0,5878+0,1050	0,5001	
1570		$5(1570 - 1518))^{-1}$ $\cong 0,0011$	0,5878+0,1631	0,4719	74,932
1588*		$(4(1588 - 1570) +$	0,7510+0,0298	0,4580	
1609*	(1570; 1973]	$3(1609 - 1588) +$	0,7510+0,0647	0,4424	
1713*		$2(1713 - 1609) +$	0,7510+0,2371	0,3723	
1973		$1(1973 - 1713))^{-1}$ $\cong 0,0017$	0,7510+0,6683	0,2419	137,671

$$\sum(\text{dias}) \cong 187,9$$

(#) mediante o cálculo dos componentes do somatório em (2.15)

Tabela A9 Estimativas (sv) da curva de sobrevivência e da correspondente média para os dados de transplante de rim, com doador vivo e tratamento A, para $a(t)=1-0,0005 t$.

j	t	V(t)	$S_T^{sv}(t)$	j_s	p_j^{sv}	$\prod p_j^{sv}$	$Sl^{sv}(t)$	sub-int _j	área _j ^{sv}
	0	25	1,0000	-	1,0	1,0	1,0000	0- 40	
	39						0,9992		39,9654
1	40	24	0,9608	-	1,0	1,0	0,9608	40- 43	
	42						0,9607		2,8630
2	43	23	0,9222	-	1,0	1,0	0,9222	43- 89	
	88						0,9214		42,3839
3	89	22	0,8829	-	1,0	1,0	0,8829	89- 201	
	200						0,8808		98,7454
4	201	21	0,8423	-	1,0	1,0	0,8423	201- 215	
	214						0,8420		11,7709
5	215	20	0,8036	-	1,0	1,0	0,8036	215- 308	
	307						0,8018		74,6285
6	308	19	0,7633	-	1,0	1,0	0,7633	308- 396	
	395						0,7616		67,0774
7	396	18	0,7231	-	1,0	1,0	0,7231	396- 512	
	511						0,7209		83,7372
8	512	17	0,6825	-	1,0	1,0	0,6825	512- 542	
	541						0,6819		20,4460
9	542	16	0,6434	-	1,0	1,0	0,6434	542- 956	264,7290
10	956*	15	0,5970	10	1,0644	1,0644	0,6355	956- 970	8,8945
11	970*	14	0,5583	11	1,0689	1,1378	0,6352	970-1227	162,4960
	1226						0,6296		
12	1227	13	0,5149	-	1,0	1,1378	0,5858	1227-1272	26,3385
13	1272*	12	0,4755	13	1,0809	1,2298	0,5848	1272-1343	41,4617
14	1343*	11	0,4357	14	1,0883	1,3383	0,5831	1343-1362	11,0748
15	1362*	10	0,3969	15	1,0969	1,4680	0,5826	1362-1424	36,0411
	1423						0,5809		
16	1424	9	0,3572	-	1,0	1,4680	0,5244	1424-1426	1,0488
17	1426*	8	0,3187	17	1,1207	1,6452	0,5244	1426-1434	4,1939
18	1434*	7	0,2801	18	1,1373	1,8711	0,5241	1434-1483	25,6385
19	1483*	6	0,2407	19	1,1598	2,1700	0,5223	1483-1518	18,2568
20	1518*	5	0,2016	20	1,1908	2,5841	0,5209	1518-1570	26,9695
	1569						0,5184		
21	1570	4	0,1621	-	1,0	2,5841	0,4189	1570-1588	7,5325
22	1588*	3	0,1233	22	1,3119	3,3901	0,4180	1588-1609	8,7642
23	1609*	2	0,0844	23	1,4555	4,9342	0,4167	1609-1713	43,8192
24	1713*	1	0,0440	24	1,8745	9,2492	0,4068	1713-1973	99,5752
	1972						0,3607		
25	1973	0	0,0005	-	1,0	9,2492	0,0048	Σ=1227,4	

- i) $S^{sv}_T(.)$ refere-se às estimativas (sv) da sobrevivência "bruta", que^T são dadas pelo primeiro termo do segundo membro de (3.13);
- ii) j_s são os valores dos índices j nos pontos de censura;
- ii) os p^{sv}_j são os termos do produto em (3.13);
- iii) Σ é a soma das áreas e portanto a estimativa da média;
- iv) as $Sl^{sv}(.)$ são dadas pelo produto da quarta e sétima colunas.

Tabela A10 Estimativas (sp) da curva de sobrevivência e da correspondente média para os dados de transplante de rim, com doador vivo tratamento A, para uma priori não informativa uniforme para ρ^* .

j	t	$S_T^{SP}(t)$	$Y_E^{SP}(t)$	$\prod^{SP}(t)$	$Sl^{SP}(t)$	sub-int _j	área _j ^{SP}
1	0	1,0000	1,0000	1,0000	1,0000	0- 40	39,966
	39	0,9986	1,0007	1,0000	0,8883		
	40	0,9615	1,0007	1,0000	0,9622	40- 43	
	42	0,9614	1,0008	1,0000	0,9622		
2	43	0,9243	1,0008	1,0000	0,9251	43- 89	42,515
	88	0,9227	1,0017	1,0000	0,9242		
3	89	0,8856	1,0017	1,0000	0,8871	89- 201	99,222
	200	0,8815	1,0041	1,0000	0,8851		
4	201	0,8444	1,0041	1,0000	0,8478	201- 215	11,849
	214	0,8439	1,0044	1,0000	0,8476		
5	215	0,8068	1,0044	1,0000	0,8104	215- 308	75,267
	307	0,8034	1,0065	1,0000	0,8087		
6	308	0,7664	1,0065	1,0000	0,7714	308- 396	67,790
	395	0,7631	1,0087	1,0000	0,7698		
7	396	0,7261	1,0087	1,0000	0,7324	396- 512	84,811
	511	0,7218	1,0117	1,0000	0,7302		
8	512	0,6847	1,0117	1,0000	0,6927	512- 542	20,755
	541	0,6837	1,0125	1,0000	0,6922		
9	542	0,6466	1,0125	1,0000	0,6547	542- 956	269,432
	956*	0,5942	1,0247	1,0623	0,6469	956- 970	
10	970*	0,5567	1,0252	1,1330	0,6466	970-1227	165,439
	1226	0,5472	1,0340	1,1330	0,6415		
11	1227	0,5101	1,0340	1,1330	0,5976	1227-1272	26,872
	1272*	0,4714	1,0357	1,2220	0,5967	1272-1343	
12	1343*	0,4317	1,0386	1,3269	0,5950	1343-1362	11,300
	1362*	0,3940	1,0395	1,4516	0,5945	1362-1424	
13	1423	0,3917	1,0425	1,4516	0,5928		1,073
	1424	0,3547	1,0425	1,4516	0,5367	1424-1426	
14	1426*	0,3176	1,0425	1,6209	0,5367	1426-1434	4,292
	1434*	0,2802	1,0426	1,8351	0,5364	1434-1483	
15	1483*	0,2414	1,0431	2,1167	0,5347	1403-1518	18,688
	1518*	0,2030	1,0465	2,5028	0,5332	1518-1570	
16	1569	0,2011	1,0494	2,5028	0,5308		7,785
	1570	0,1641	1,0543	2,5028	0,4330	1570-1588	
17	1588*	0,1264	1,0544	3,2364	0,4321	1588-1609	9,060
	1609*	0,0886	1,0565	4,5899	0,4308	1609-1713	
18	1713*	0,0477	1,0837	8,1563	0,4213	1713-1973	103,624
	1972	0,0010	1,2125	8,1563	0,3765		
19	1973	0,0010	1,2168	8,1563	0,0099		Σ=1248,9

Obs: * $\rho=P[Te<To]$

Tabela A11 Cálculo da estimativa (mk), da curva de sobrevivência para os dados de transplante de rim, doador vivo, esquema A.

t	(t1 _m -t1 _{m-1}]	Δt1 _m	M[t1 _m]	W[t1 _m]	V[t1 _m ⁻]	mk r1	mk R1(t)=[R1 ^{mk} +(t-t1 _{m-1})r1]	mk S1(t)
0								1,0000
40	(0; 40]	40	1	0	25	0,0010	0,0 +0,0400	0,9608
43	(40; 43]	3	1	0	24	0,0139	0,0400+0,0417	0,9216
89	(43; 89]	46	1	0	23	0,0009	0,0817+0,0435	0,8824
201	(89; 201]	112	1	0	22	0,0004	0,1251+0,0454	0,8432
215	(201; 215]	14	1	0	21	0,0034	0,1706+0,0476	0,8039
308	(215; 308]	93	1	0	20	0,0005	0,2182+0,0500	0,7647
396	(308; 396]	88	1	0	19	0,0006	0,2682+0,0526	0,7255
512	(396; 512]	116	1	0	18	0,0005	0,3208+0,0556	0,6863
542	(512; 542]	30	1	0	17	0,0020	0,3764+0,0588	0,6471
956*							0,4352+0,0417	0,6207
970*							0,4352+0,0431	0,6198
1227	(542;1227]	685	1	2	14	0,0001	0,4352+0,0690	0,6039
1272*							0,5043+0,0213	0,5912
1343*							0,5043+0,0550	0,5716
1362*							0,5043+0,0641	0,5665
1424	(1227;1424]	197	1	3	10	0,0005	0,5043+0,0935	0,5500
1426*							0,5977+0,0023	0,5487
1434*							0,5977+0,0117	0,5436
1483*							0,5977+0,0693	0,5132
1518*							0,5977+0,1104	0,4926
1570	(1424;1570]	146	1	4	5	0,0012	0,5977+0,1714	0,4634
1588*							0,7692+0,0313	0,4491
1609*							0,7692+0,0677	0,4330
1713*							0,7692+0,2484	0,3615
1973	(1570;1973]	403	1	3	1	0,0017	0,7692+0,7000	0,2301

1

1

mk

V[t1_m⁻]

Δt1

(V[t1_m⁻]+W[t1_m]/2)

Δt1

2

(*)

r1(t1_m)=

mk

R1(t)=

∑

t1_{m-1}<t

[t1_{m-1}-t1_{m-2}]

r1(t1_{m-1})

mk

, com t∈(t1_{m-1},t1_m]

mk

S1(t)=exp(-R1(t))

(***)

150