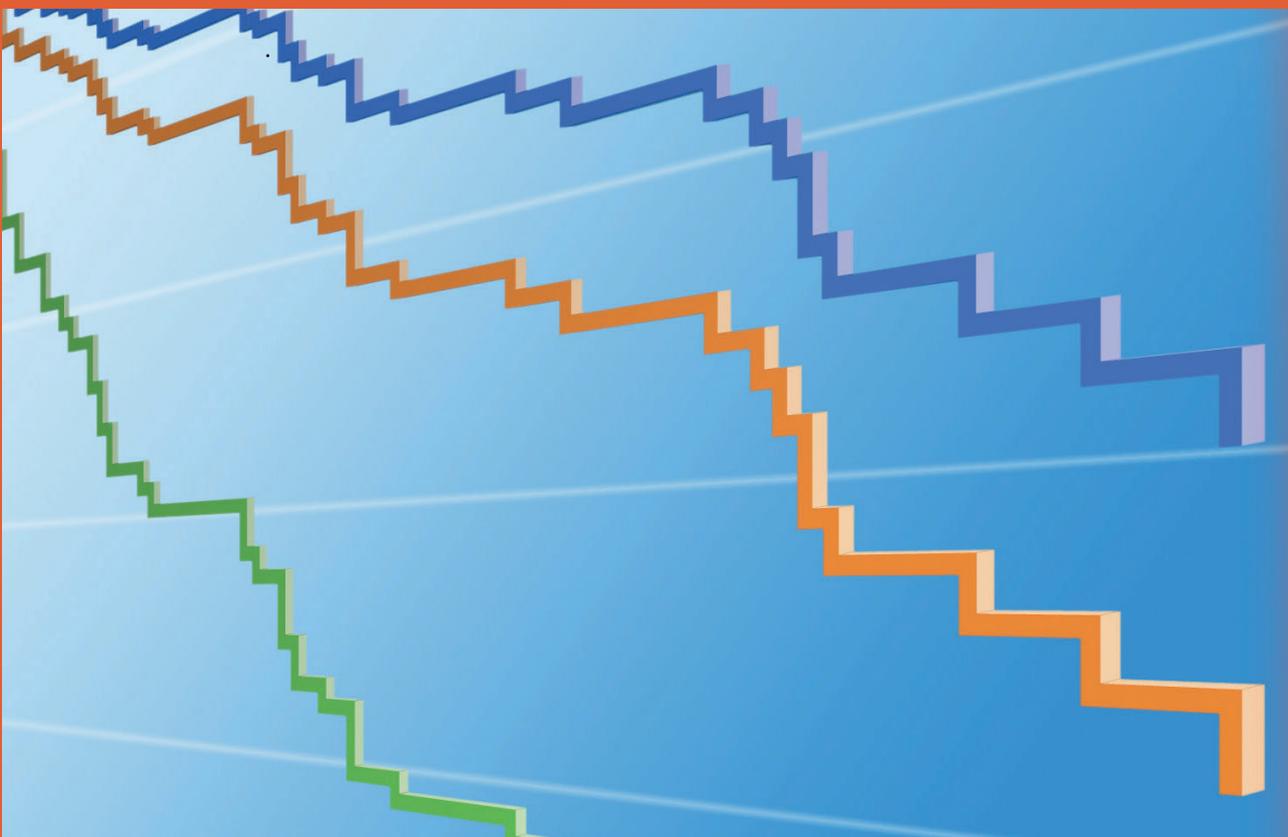
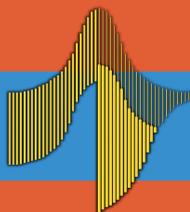


Enrico Antônio Colosimo
Suely Ruiz Giolo

ANÁLISE DE SOBREVIVÊNCIA APLICADA



Blucher



ABE - PROJETO FISHER



Enrico Antônio Colosimo
Departamento de Estatística, UFMG

Suely Ruiz Giolo
Departamento de Estatística, UFPR



ABE - PROJETO FISHER

Análise de sobrevivência aplicada

© 2006 Enrico Antônio Colosimo

Suely Ruiz Giolo

1ª edição – 2006

3ª reimpressão – 2019

Editora Edgard Blücher Ltda.

Blucher

Rua Pedroso Alvarenga, 1245, 4º andar

04531-012 – São Paulo – SP – Brasil

Tel 55 11 3078-5366

contato@blucher.com.br

www.blucher.com.br

É proibida a reprodução total ou parcial por quaisquer meios, sem autorização escrita da editora.

Todos os direitos reservados pela Editora Edgard Blücher Ltda.

FICHA CATALOGRÁFICA

Colosimo, Enrico Antônio

Análise de sobrevivência aplicada / Enrico Antônio Colosimo, Suely Ruiz Giolo. – São Paulo: Blucher, 2006.

Bibliografia.

ISBN 978-85-212-0384-1

1. Análise de sobrevivência (Biometria)
2. Estatística I. Giolo, Suely Ruiz. II. Título.

05-8936

CDD-519.5

Índice para catálogo sistemático:

1. Análise de sobrevivência: Estatística aplicada 519.5

Conteúdo

Prefácio	xiii
1 Conceitos Básicos e Exemplos	1
1.1 Introdução	1
1.2 Objetivo e Planejamento dos Estudos	3
1.3 Caracterizando Dados de Sobrevivência	6
1.3.1 Tempo de Falha	7
1.3.2 Censura e Dados Truncados	8
1.4 Representação dos Dados de Sobrevivência	12
1.5 Exemplos de Dados de Sobrevivência	13
1.5.1 Dados de Hepatite	13
1.5.2 Dados de Malária	14
1.5.3 Dados de Leucemia Pediátrica	15
1.5.4 Dados de Sinusite em Pacientes Infectados pelo HIV	16
1.5.5 Dados de Aleitamento Materno	18
1.5.6 Dados Experimentais utilizando Camundongos	18
1.5.7 Dados de Câncer de Mama	19
1.5.8 Dados de Tempo de Vida de Mangueiras	19
1.6 Especificando o Tempo de Sobrevivência	20
1.6.1 Função de Sobrevivência	20
1.6.2 Função de Taxa de Falha	22
1.6.3 Função de Taxa de Falha Acumulada	23

1.6.4	Tempo Médio e Vida Média Residual	24
1.6.5	Relações entre as Funções	25
1.7	Exercícios	26
2	Técnicas Não-Paramétricas	29
2.1	Introdução	29
2.2	Estimação na Ausência de Censura	32
2.3	O Estimador de Kaplan-Meier	34
2.4	Outros Estimadores Não-Paramétricos	42
2.4.1	Estimador de Nelson-Aalen	43
2.4.2	Estimador da Tabela de Vida ou Atuarial	45
2.4.3	Comparação dos Estimadores de $S(t)$	47
2.5	Estimação de Quantidades Básicas	48
2.5.1	Exemplo: Reincidência de Tumor Sólido	51
2.6	Comparação de Curvas de Sobrevivência	55
2.6.1	Análise dos Dados da Malária	59
2.6.2	Outros Testes	61
2.7	Exercícios	63
3	Modelos Probabilísticos	69
3.1	Introdução	69
3.2	Modelos em Análise de Sobrevivência	70
3.2.1	Distribuição Exponencial	70
3.2.2	Distribuição de Weibull	72
3.2.3	Distribuição Log-normal	75
3.2.4	Distribuição Log-logística	77
3.2.5	Distribuições Gama e Gama Generalizada	79
3.2.6	Outros Modelos Probabilísticos	81
3.3	Estimação dos Parâmetros dos Modelos	83
3.3.1	O Método de Máxima Verossimilhança	84
3.3.2	Exemplos de Aplicações	87

3.4	Intervalos de Confiança e Testes de Hipóteses	90
3.4.1	Intervalos de Confiança	91
3.4.2	Testes de Hipóteses	93
3.5	Escolha do Modelo Probabilístico	95
3.5.1	Métodos Gráficos	96
3.5.2	Comparação de Modelos	100
3.6	Exemplos	101
3.6.1	Exemplo 1 - Pacientes com Câncer de Bexiga	101
3.6.2	Exemplo 2 - Tratamento Quimioterápico	108
3.7	Exercícios	111
4	Modelos de Regressão Paramétricos	115
4.1	Introdução	115
4.2	Modelo Linear para Dados de Sobrevivência	116
4.2.1	Modelo de Regressão Exponencial	118
4.2.2	Modelo de Regressão Weibull	121
4.2.3	Modelo de Tempo de Vida Acelerado	122
4.3	Adequação do Modelo Ajustado	123
4.3.1	Resíduos de Cox-Snell	124
4.3.2	Resíduos Padronizados	125
4.3.3	Resíduos Martingal	126
4.3.4	Resíduos Deviance	127
4.4	Interpretação dos Coeficientes Estimados	127
4.5	Exemplos	129
4.5.1	Sobrevida de Pacientes com Leucemia Aguda	129
4.5.2	Grupos de Pacientes com Leucemia Aguda	134
4.5.3	Análise dos Dados de Aleitamento Materno	140
4.6	Exercícios	153
5	Modelo de Regressão de Cox	155
5.1	Introdução	155

- 5.2 O Modelo de Cox 156
- 5.3 Ajustando o Modelo de Cox 159
 - 5.3.1 Método de Máxima Verossimilhança Parcial 160
- 5.4 Interpretação dos Coeficientes 162
- 5.5 Estimando Funções Relacionadas a $\lambda_0(t)$ 164
- 5.6 Adequação do Modelo de Cox 165
 - 5.6.1 Avaliação da Qualidade Geral do Modelo Ajustado . 166
 - 5.6.2 Avaliação da Proporcionalidade 166
 - 5.6.3 Avaliação de outros Aspectos do Modelo de Cox . . 172
- 5.7 Exemplos 175
 - 5.7.1 Análise de um Estudo sobre Câncer de Laringe . . . 175
 - 5.7.2 Análise dos Dados de Aleitamento Materno 183
 - 5.7.3 Análise dos Dados de Leucemia Pediátrica 190
- 5.8 Comentários sobre o Modelo de Cox 198
- 5.9 Exercícios 199

- 6 Extensões do Modelo de Cox 201**
 - 6.1 Introdução 201
 - 6.2 Modelo de Cox com Covariáveis Dependentes do Tempo . . 202
 - 6.3 Modelo de Cox Estratificado 204
 - 6.4 Análise dos Dados de Pacientes HIV 206
 - 6.4.1 Descrição dos Dados 206
 - 6.4.2 Modelagem Estatística 208
 - 6.5 Modelo de Cox Estratificado nos Dados de Leucemia 210
 - 6.6 Estudo sobre Hormônio de Crescimento 215
 - 6.6.1 Resultados do Modelo de Cox Estratificado 219
 - 6.7 Exercícios 223

- 7 Modelo Aditivo de Aalen 225**
 - 7.1 Introdução 225
 - 7.2 Modelo Aditivo de Aalen 227

7.3	Estimação	228
7.4	Teste para os Efeitos das Covariáveis	232
7.5	Diagnóstico do Modelo	234
7.6	Análise dos Dados de Câncer de Laringe	234
7.7	Análise dos Dados de Pacientes com HIV	239
7.7.1	Considerações Finais	244
7.8	Exercícios	244
8	Censura Intervalar e Dados Grupados	245
8.1	Introdução	245
8.2	Técnicas Não-Paramétricas	248
8.2.1	Exemplo de Câncer de Mama	250
8.3	Modelos Paramétricos	254
8.3.1	Análise dos Dados de Câncer de Mama	255
8.4	Modelo Semiparamétrico	258
8.4.1	Modelo de Cox para os Dados de Câncer de Mama	260
8.5	Dados Grupados	264
8.6	Aproximações para a Verossimilhança Parcial	264
8.7	Modelos de Regressão Discretos	268
8.7.1	Modelo de Taxas de Falha Proporcionais	269
8.7.2	Modelo Logístico	270
8.8	Aplicação: Ensaio de Vida de Mangueiras	271
8.9	Modelos Discretos ou Aproximações?	278
8.10	Exercícios	280
9	Análise de Sobrevida Multivariada	281
9.1	Introdução	281
9.2	Fragilidade em um Contexto Univariado	283
9.3	Fragilidade em um Contexto Multivariado	285
9.3.1	Modelo de Fragilidade Compartilhado	285
9.4	Generalizações do Modelo de Fragilidade	287

- 9.4.1 Modelo de Fragilidade Estratificado 287
- 9.4.2 Modelo de Fragilidade com Associações Complexas . 288
- 9.4.3 Modelo de Fragilidade Multiplicativo 288
- 9.4.4 Modelo de Fragilidade Aditivo 288
- 9.4.5 Modelo de Fragilidade Dependente do Tempo 290
- 9.5 Distribuições para a Variável de Fragilidade 290
- 9.6 Modelo de Fragilidade Gama 292
- 9.7 Estimaco no Modelo de Fragilidade Gama 293
 - 9.7.1 Estimaco via Algoritmo EM 294
 - 9.7.2 Estimaco via Verossimilhana Penalizada 299
 - 9.7.3 Estimaco Bayesiana via MCMC 301
- 9.8 Testando a Fragilidade 301
 - 9.8.1 Testando o Efeito das Covariáveis 303
- 9.9 Diagnstico dos Modelos de Fragilidade 303
- 9.10 Modelando Eventos Recorrentes 304
 - 9.10.1 Formulaco de Andersen e Gill (AG) 305
 - 9.10.2 Formulaco de Wei, Lin e Weissfeld (WLW) 307
 - 9.10.3 Formulaco de Prentice, Williams e Peterson (PWP) 308
 - 9.10.4 Consideraco sobre os Modelos Marginais 309
- 9.11 Exemplos 310
 - 9.11.1 Fragilidade no Estudo de Leucemia Pediátrica 310
 - 9.11.2 Estudo com Animais da Raa Nelore 312
- 9.12 Exercícios 317

Apêndice 319

Referências Bibliográficas 355

Capítulo 1

Conceitos Básicos e Exemplos

1.1 Introdução

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas duas décadas do século passado. A razão deste crescimento é o desenvolvimento e aprimoramento de técnicas estatísticas combinado com computadores cada vez mais velozes. Uma evidência quantitativa deste sucesso é o número de aplicações de análise de sobrevivência em medicina. Bailar III e Mosteller (1992, Capítulo 3) verificaram que o uso de métodos de análise de sobrevivência cresceu de 11%, em 1979, para 32%, em 1989, nos artigos do conceituado periódico *The New England Journal of Medicine*. Esta foi a área da estatística, segundo os autores, que mais se destacou no período avaliado. Os dois artigos mais citados em toda a literatura estatística no período de 1987 a 1989 foram, segundo Stigler (1994), o do estimador de Kaplan-Meier para a função de sobrevivência (Kaplan e Meier, 1958) e o do modelo de Cox (Cox, 1972).

Em análise de sobrevivência, a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Este tempo é denominado **tempo de falha**, podendo ser o tempo até a morte do paciente, bem como até a cura ou recidiva de uma doença. Em estudos de câncer, é usual o registro das datas correspondentes ao diagnóstico da doença, à

remissão (após o tratamento, o paciente fica livre dos sintomas da doença), à recorrência da doença (recidiva) e à morte do paciente. O tempo de falha pode ser, por exemplo, do diagnóstico até a morte ou da remissão até a recidiva.

A principal característica de dados de sobrevivência é a presença de **censura**, que é a observação parcial da resposta. Isto se refere a situações em que, por alguma razão, o acompanhamento do paciente foi interrompido, seja porque o paciente mudou de cidade, o estudo terminou para a análise dos dados ou, o paciente morreu de causa diferente da estudada. Isto significa que toda informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado. Sem a presença de censura, as técnicas estatísticas clássicas, como análise de regressão e planejamento de experimentos, poderiam ser utilizadas na análise deste tipo de dados, provavelmente usando uma transformação para a resposta. Suponha, por exemplo, que o interesse seja o de comparar o tempo médio de vida de três grupos de pacientes. Se não houver censuras, pode-se usar as técnicas usuais de análise de variância para se fazer tal comparação. No entanto, se houver censuras, que é o mais provável, tais técnicas não podem ser utilizadas pois elas necessitam de todos os tempos de falha. Desta forma, faz-se necessário o uso dos métodos de análise de sobrevivência que possibilitam incorporar na análise estatística a informação contida nos dados censurados.

O termo análise de sobrevivência refere-se basicamente a situações médicas envolvendo dados censurados. Entretanto, condições similares ocorrem em outras áreas em que se usam as mesmas técnicas de análise de dados. Em engenharia, são comuns os estudos em que produtos ou componentes são colocados sob teste para se estimar características relacionadas aos seus tempos de vida, tais como o tempo médio ou a probabilidade de um certo produto durar mais do que 5 anos. Exemplos podem ser encontrados em Nelson (1990a), Meeker e Escobar (1998) e Freitas e Colosimo (1997). Os

engenheiros denominam esta área de confiabilidade. O mesmo ocorre em ciências sociais, em que várias situações de interesse têm como resposta o tempo entre eventos (Allison, 1984; Elandt-Johnson e Johnson, 1980). Criminalistas estudam o tempo entre a liberação de presos e a ocorrência de crimes; estudiosos do trabalho se concentram em mudanças de empregos, desempregos, promoções e aposentadorias; demógrafos, com nascimentos, mortes, casamentos, divórcios e migrações. O crescimento observado no número de aplicações em medicina também pode ser observado nessas outras áreas.

Este texto foi motivado por ilustrações essencialmente na área clínica. Desta forma, os exemplos e colocações são conduzidos para esta área. No entanto, enfatiza-se que as técnicas estatísticas são de ampla utilização em outras áreas do conhecimento, como mencionado anteriormente.

Este capítulo é dedicado à apresentação de conceitos básicos e definições de funções importantes para a análise de dados de sobrevivência. Os objetivos e planejamento de alguns estudos clínicos e industriais são discutidos na Seção 1.2. A caracterização e representação dos dados de sobrevivência são apresentadas nas Seções 1.3 e 1.4. Vários exemplos de aplicação das técnicas de análise de sobrevivência são descritos na Seção 1.5. A Seção 1.6 finaliza o capítulo apresentando as principais funções de interesse na análise de dados de sobrevivência, bem como algumas relações matemáticas importantes entre elas.

1.2 Objetivo e Planejamento dos Estudos

Os estudos clínicos são investigações científicas realizadas com o objetivo de verificar uma determinada hipótese de interesse. Estas investigações são conduzidas coletando dados e analisando-os por meio de métodos estatísticos. Em geral, estes estudos podem ser divididos nas seguintes três etapas:

1. Formulação da hipótese de interesse.
2. Planejamento e coleta dos dados.
3. Análise estatística dos dados para testar a hipótese formulada.

Estas etapas são comuns em qualquer estudo envolvendo análise estatística de dados. No presente texto, o interesse envolve situações em que a variável resposta é o tempo até a ocorrência de um evento de interesse, como descrito na Seção 1.1.

A primeira etapa de um estudo clínico é gerada pela curiosidade científica do pesquisador. Identificar fatores de risco ou prognóstico para uma doença é um objetivo que aparece com frequência em estudos clínicos. A comparação de drogas ou diferentes opções terapêuticas é outro objetivo usualmente encontrado neste tipo de estudo.

Os textos técnicos estatísticos concentram todo o esforço na terceira etapa, ou seja, na análise estatística dos dados, mesmo admitindo a importância de um adequado planejamento do estudo. Este texto não é diferente dos demais. No entanto, o restante desta seção é dedicado a uma breve descrição desta segunda etapa.

Em análise de sobrevivência, a resposta é por natureza longitudinal. O delineamento de estudos com respostas dessa natureza pode ser observacional ou experimental, assim como ele pode ser retrospectivo ou prospectivo. As quatro formas básicas de estudos clínicos são: descritivo, caso-controle, coorte e clínico aleatorizado. Os três primeiros são observacionais e o quarto é experimental, pois existe a intervenção do pesquisador ao alocar, de forma aleatória, tratamento ao paciente. O uso das técnicas de análise de sobrevivência é mais frequente nos estudos de coorte e ensaios clínicos. Entretanto, o seu uso é também possível nos demais estudos, desde que os tempos até a ocorrência do evento de interesse possam ser claramente definidos e obtidos.

O estudo envolvendo somente uma amostra, usualmente de doentes, é descritivo, pois não existe um grupo de comparação. Nestes estudos o objetivo é freqüentemente a identificação de fatores de prognóstico para a doença em estudo. Os outros tipos de estudo são comparativos. Isto significa que o objetivo do estudo é a comparação de dois ou mais grupos.

O estudo caso-controle é usualmente retrospectivo. Dois grupos, um de doentes (casos) e outro de não-doentes (controles), são comparados em relação à exposição a um ou mais fatores de interesse. Este estudo é simples, de baixo custo e rápido, pois a informação já se encontra disponível. No entanto, ele sofre algumas limitações por estar sujeito a alguns tipos de vícios. Esses vícios estão relacionados à informação disponível sobre a história da exposição, assim como a incerteza sobre a escolha do grupo controle. Uma discussão mais profunda sobre este tipo de estudo foge do escopo deste livro. No entanto, devido a sua grande utilização, existem várias bibliografias sobre o assunto, entre elas, Breslow e Day (1980) e Rothman e Greenland (1998).

As limitações dos estudos caso-controle podem ser vencidas pelos estudos conhecidos por coorte. Em um estudo de coorte, dois grupos, um exposto e outro não-exposto ao fator de interesse, são acompanhados por um período de tempo registrando-se a ocorrência da doença ou evento de interesse. A vantagem deste estudo sobre o caso-controle é poder avaliar a comparabilidade dos grupos no início do estudo e identificar as variáveis de interesse a serem medidas. Por outro lado, é um estudo longo e mais caro, pois os indivíduos são acompanhados por um período de tempo muitas vezes superior a um ano. Também não é um estudo indicado para doenças consideradas raras. Uma referência importante sobre esses estudos é Breslow e Day (1987).

A forma mais consagrada de pesquisa clínica é o estudo clínico aleatorizado, que é importante por ser experimental. Isto significa que existe a intervenção direta do pesquisador ao alocar, de forma aleatória, tratamento

ao paciente. Este procedimento garante a comparabilidade dos grupos. Este estudo é bastante analisado na literatura e pode-se citar os seguintes livros, entre outros, Pocock (1983) e Friedman et al. (1998).

Na Seção 1.5 são descritos alguns exemplos reais analisados ao longo do texto. Dentre eles, existem estudos descritivos, de coorte e clínico aleatorizado.

Os estudos industriais são usualmente de campo ou realizados na própria empresa simulando situações de campo. No entanto, existem alguns estudos industriais planejados com o objetivo de reduzir o tempo de vida das unidades sob teste e, desta forma, obter dados amostrais mais rápidos. Esses estudos são denominados de testes de vida acelerados. Os itens amostrais são estressados para falhar mais rápido e por meio de modelos de regressão obtêm-se as estimativas para as quantidades de interesse nas condições de uso utilizando extrapolações. Uma discussão mais profunda sobre estes testes pode ser encontrada em Nelson (1990a) e Freitas e Colosimo (1997).

Uma extensão destes testes é o de degradação, que pode ser ou não acelerado. Nestes testes, uma variável numérica associada ao tempo de falha é registrada ao longo do período de acompanhamento. A partir dos valores desta variável, é possível obter as estimativas de interesse mesmo em situações em que nenhuma falha tenha sido registrada. Estes testes estão ganhando espaço na literatura técnica de engenharia. Mais informações sobre eles podem ser encontradas em Meeker e Escobar (1998) e Oliveira e Colosimo (2004).

1.3 Caracterizando Dados de Sobrevida

Os conjuntos de dados de sobrevivência são caracterizados pelos tempos de falha e, muito freqüentemente, pelas censuras. Estes dois componentes constituem a resposta. Em estudos clínicos, um conjunto de covariáveis é também, geralmente, medido em cada paciente. Os seguintes elementos

constituem o tempo de falha: o tempo inicial, a escala de medida e o evento de interesse (falha). Estes três elementos devem ser claramente definidos e, juntamente com a censura, são discutidos em detalhes a seguir.

1.3.1 Tempo de Falha

O tempo de início do estudo deve ser precisamente definido. Os indivíduos devem ser comparáveis na origem do estudo, com exceção de diferenças medidas pelas covariáveis. Em um estudo clínico aleatorizado, a data da aleatorização é a escolha natural para a origem do estudo. A data do início do tratamento de doenças ou do diagnóstico também são outras escolhas possíveis.

A escala de medida é quase sempre o tempo real ou “de relógio”, apesar de existirem outras alternativas. Em testes de engenharia podem surgir outras escalas de medida, como o número de ciclos, a quilometragem de um carro ou qualquer outra medida de carga.

O terceiro elemento é o evento de interesse. Estes eventos são, na maioria dos casos, indesejáveis e, como já mencionado, chamados de falha. É importante, em estudos de sobrevivência, definir de forma clara e precisa o que vem a ser a falha. Em algumas situações, a definição de falha já é clara, tais como morte ou recidiva, mas em outras pode assumir termos ambíguos. Por exemplo, fabricantes de produtos alimentícios desejam saber informações sobre o tempo de vida de seus produtos expostos em balcões frigoríficos de supermercados. O tempo de falha vai do tempo inicial de exposição (chegada ao supermercado) até o produto ficar “inapropriado ao consumo”. Este evento deve ser claramente definido antes do estudo ter seu início. Por exemplo, o produto fica inadequado para o consumo quando atingir mais do que uma determinada concentração de microorganismos por mm^2 de área do produto.

O evento de interesse (falha) pode ainda ocorrer devido a uma única causa ou devido a duas ou mais. Situações em que causas de falha com-

petem entre si são denominadas na literatura de *riscos competitivos* (Prentice et al., 1978). Apenas as que consideram uma única causa de falha são abordadas neste texto.

1.3.2 Censura e Dados Truncados

Os estudos clínicos que envolvem uma resposta temporal são frequentemente prospectivos e de longa duração. Mesmo sendo longos, os estudos clínicos de sobrevivência usualmente terminam antes que todos os indivíduos no estudo venham a falhar. Uma característica decorrente destes estudos é, então, a presença de observações incompletas ou parciais. Estas observações, denominadas censuras, podem ocorrer por uma variedade de razões, dentre elas, a perda de acompanhamento do paciente no decorrer do estudo e a não ocorrência do evento de interesse até o término do experimento. Note que toda informação obtida sobre estes indivíduos é que o tempo até a ocorrência do evento, para cada um deles, é superior ao tempo registrado até o último acompanhamento.

Ressalta-se o fato de que, mesmo censurados, todos os resultados provenientes de um estudo de sobrevivência devem ser usados na análise estatística. Duas razões justificam tal procedimento: (i) mesmo sendo incompletas, as observações censuradas fornecem informações sobre o tempo de vida de pacientes; (ii) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar conclusões viciadas.

Alguns mecanismos de censura são diferenciados em estudos clínicos. Censuras do tipo I ocorrem naqueles estudos que ao serem finalizados após um período pré-estabelecido de tempo registram, em seu término, alguns indivíduos que ainda não apresentaram o evento de interesse. Censuras do tipo II resultam de estudos os quais são finalizados após a ocorrência do evento de interesse em um número pré-estabelecido de indivíduos. Um terceiro mecanismo de censura, o do tipo aleatório, é o que mais ocorre na prática médica. Isto acontece quando um paciente é retirado no decorrer

do estudo sem ter ocorrido a falha, ou também, por exemplo, se o paciente morrer por uma razão diferente da estudada.

Uma representação simples do mecanismo de censura aleatória é feita usando duas variáveis aleatórias. Considere T uma variável aleatória representando o tempo de falha de um paciente e C , uma outra variável aleatória independente de T , representando o tempo de censura associado a este paciente. O que se observa para este paciente é, portanto,

$$t = \min(T, C)$$

$$e \quad \delta = \begin{cases} 1 & \text{se } T \leq C \\ 0 & \text{se } T > C. \end{cases}$$

Suponha que os pares (T_i, C_i) , para $i = 1, \dots, n$, formam uma amostra aleatória de n pacientes. Observe que se todo $C_i = C$, uma constante fixa sob o controle do pesquisador, tem-se a censura do tipo I. Ou seja, a censura do tipo I é um caso particular da aleatória. Observe que neste caso, a variável aleatória t tem uma probabilidade maior do que zero em $t = C$. Isto significa que, no caso de censura do tipo I, t é uma variável aleatória mista com um componente contínuo e outro discreto. A Figura 1.1 ilustra os mecanismos de censura descritos.

Os mecanismos de censura apresentados na Figura 1.1 são conhecidos por censura à direita, pois o tempo de ocorrência do evento de interesse está à direita do tempo registrado. Esta é a situação freqüentemente encontrada em estudos envolvendo dados de sobrevivência. No entanto, outras duas formas de censura podem ocorrer: censura à esquerda e intervalar.

A censura à esquerda ocorre quando o tempo registrado é maior do que o tempo de falha. Isto é, o evento de interesse já aconteceu quando o indivíduo foi observado. Um estudo para determinar a idade em que as crianças aprendem a ler em uma determinada comunidade pode ilustrar a situação de censura à esquerda. Quando os pesquisadores começaram a pesquisa algumas crianças já sabiam ler e não lembravam com que idade

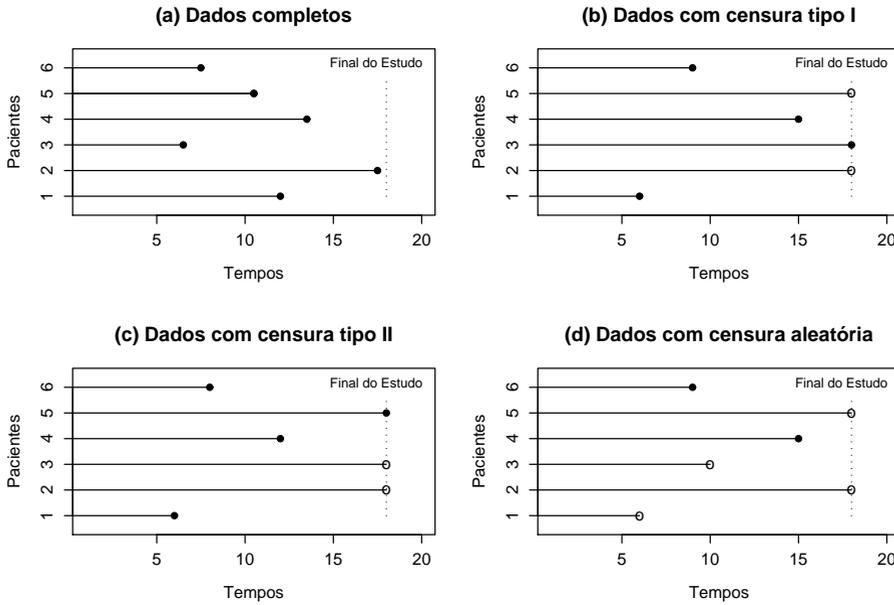


Figura 1.1: Ilustração de alguns mecanismos de censura em que ● representa a falha e ○ a censura. (a) todos os pacientes experimentaram o evento antes do final do estudo, (b) alguns pacientes não experimentaram o evento até o final do estudo, (c) o estudo foi finalizado após a ocorrência de um número pré-estabelecido de falhas e (d) o acompanhamento de alguns pacientes foi interrompido por alguma razão e alguns pacientes não experimentaram o evento até o final do estudo.

isto tinha acontecido, caracterizando, desta forma, observações censuradas à esquerda. Neste mesmo estudo, pode ocorrer simultaneamente censura à direita para crianças que não sabiam ler quando os dados foram coletados. Os tempos de vida neste caso são chamados de duplamente censurados (Turnbull, 1974).

A intervalar é um tipo mais geral de censura que acontece, por exemplo, em estudos em que os pacientes são acompanhados em visitas periódicas e é conhecido somente que o evento de interesse ocorreu em um certo intervalo de tempo. Pelo fato de o tempo de falha T não ser conhecido exatamente, mas sim pertencer a um intervalo, isto é, $T \in (L, U]$, tais dados são denominados por *sobrevivência intervalar* ou, mais usualmente, por *dados de*

censura intervalar. Lindsey et al. (1998) observam que tempos exatos de falha, bem como censuras à direita e à esquerda, são casos especiais de dados de sobrevivência intervalar com $L = U$ para tempos exatos de falha, $U = \infty$ para censuras à direita e $L = 0$ para censuras à esquerda.

Uma outra característica de alguns estudos de sobrevivência é o truncamento que é muitas vezes confundido com censura. Truncamento é caracterizado por uma condição que exclui certos indivíduos do estudo. Nestes estudos, os pacientes não são acompanhados a partir do tempo inicial, mas somente após experimentarem um certo evento. Por exemplo, isto acontece se, para estimação da distribuição do tempo de vida dos moradores de uma certa localidade, for usada uma amostra retirada do banco de dados da previdência local. Desta forma, somente moradores que atingiram a aposentadoria fazem parte da amostra. Estas observações são conhecidas por truncadas à esquerda. Em estudos de AIDS, a data da infecção é uma origem de tempo bastante utilizada e o evento de interesse pode ser o desenvolvimento da AIDS. Neste caso, o número de pacientes infectados é desconhecido. Então, indivíduos já infectados e que ainda não desenvolveram a doença são desconhecidos para o pesquisador e não são incluídos na amostra. Neste caso, somente pacientes que têm comprovada a doença fazem parte da amostra. Estas observações são chamadas de truncadas à direita. Outros exemplos de truncamento podem ser encontrados em Nelson (1990b), Kalbfleisch e Lawless (1992) e Klein e Moeschberger (2003).

A presença de censuras traz problemas para a análise estatística. A censura do tipo II é, em princípio, mais tratável que os outros tipos. Métodos exatos de inferência estatística existem para a censura do tipo II, mas para situações bem simples que raramente acontecem em estudos clínicos (Lawless, 2003). Na prática, faz-se uso de resultados assintóticos para se realizar a análise estatística dos dados de sobrevivência. Esses resultados não exigem o reconhecimento do mecanismo de censura e, desse modo, as mesmas

técnicas estatísticas são utilizadas na análise de dados oriundos dos três mecanismos de censura.

Neste texto, a atenção está voltada aos dados de sobrevivência com censura à direita, que é a situação encontrada com mais frequência em estudos, tanto em medicina quanto em engenharia e ciências sociais. Um tratamento geral para dados censurados e truncados pode ser encontrado em Turnbull (1976) e Klein e Moeschberger (2003). No caso particular de dados de sobrevivência com censura intervalar, algumas técnicas especializadas de análise são apresentadas no Capítulo 8. Desta forma, quando for simplesmente mencionada a palavra *censura* entenda-se, neste texto, censura à direita.

1.4 Representação dos Dados de Sobrevivência

Os dados de sobrevivência para o indivíduo i ($i = 1, \dots, n$) sob estudo são representados, em geral, pelo par (t_i, δ_i) sendo t_i o tempo de falha ou de censura e δ_i a variável indicadora de falha ou censura, isto é,

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é um tempo de falha} \\ 0 & \text{se } t_i \text{ é um tempo censurado.} \end{cases}$$

Desta forma, a variável aleatória resposta em análise de sobrevivência é representada por duas colunas no banco de dados.

Na presença de covariáveis medidas no i -ésimo indivíduo, tais como, dentre outras, $\mathbf{x}_i = (\text{sexo}_i, \text{idade}_i, \text{tratamento}_i)$, os dados ficam representados por $(t_i, \delta_i, \mathbf{x}_i)$. No caso particular de dados de sobrevivência intervalar, tem-se, ainda, a representação $(\ell_i, u_i, \delta_i, \mathbf{x}_i)$ em que ℓ_i e u_i são, respectivamente, os limites inferior e superior do intervalo observado para o i -ésimo indivíduo.

1.5 Exemplos de Dados de Sobrevivência

Existem vários exemplos de aplicação das técnicas de análise de sobrevivência. Na área médica, elas são muito utilizadas na identificação de fatores de prognóstico para uma doença, bem como na comparação de tratamentos. Em oncologia, qualquer nova terapêutica ou droga para o combate ao câncer requer um estudo em que a resposta de interesse seja, geralmente, o tempo de sobrevivência dos pacientes. Esta resposta é denominada sobrevida global pelos oncologistas. Estudos epidemiológicos da AIDS são outros exemplos em que as técnicas de análise de sobrevivência vêm sendo usadas com frequência. Jacobson et al. (1993) mostram um estudo típico nesta área.

A seguir são descritos brevemente alguns dos exemplos utilizados no restante do texto para ilustrar as técnicas estatísticas descritas. Estes exemplos são situações reais provenientes de assessorias estatísticas dos autores, assim como alguns de literatura técnica na área médica. Os dados encontram-se no Apêndice.

1.5.1 Dados de Hepatite

Um estudo clínico aleatorizado foi realizado para investigar o efeito da terapia com esteróide no tratamento de hepatite viral aguda (Gregory et al., 1976). Vinte e nove pacientes com esta doença foram aleatorizados para receber um placebo ou o tratamento com esteróide. Cada paciente foi acompanhado por 16 semanas ou até a morte (evento de interesse) ou até a perda de acompanhamento. Os tempos de sobrevivência observados, em semanas, para os dois grupos são apresentados na Tabela 1.1. O símbolo + indica censura.

Este exemplo, que é caracterizado pela censura do tipo aleatória, é utilizado no Capítulo 2 para ilustrar as técnicas não-paramétricas para dados de sobrevivência.

Tabela 1.1: Tempos, em semanas, observados no estudo de hepatite.

Grupos	Tempos de sobrevivência
Controle	1+, 2+, 3, 3, 3+, 5+, 5+, 16+, 16+, 16+, 16+, 16+, 16+, 16+, 16+
Esteróide	1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

1.5.2 Dados de Malária

Um estudo experimental realizado com camundongos para verificar a eficácia da imunização pela malária foi conduzido no Centro de Pesquisas René Rachou, Fiocruz, MG. Nesse estudo, quarenta e quatro camundongos foram aleatoriamente divididos em três grupos e todos foram infectados pela malária (*Plasmodium berguei*). Os camundongos do grupo 1 foram imunizados 30 dias antes da infecção. Além da infecção pela malária, os camundongos dos grupos 1 e 3 foram, também, infectados pela esquistossomose (*Schistosoma mansoni*). A resposta de interesse nesse estudo foi o tempo decorrido desde a infecção pela malária até a morte do camundongo. Este tempo foi medido em dias e o estudo foi acompanhado por 30 dias. Os tempos de sobrevivência observados para os três grupos encontram-se na Tabela 1.2. O símbolo + indica censura.

Tabela 1.2: Tempos, em dias, observados no estudo da malária.

Grupos (total)	Tempos de sobrevivência
Grupo 1 (16)	7, 8, 8, 8, 8, 12, 12, 17, 18, 22, 30+, 30+, 30+, 30+, 30+, 30+
Grupo 2 (15)	8, 8, 9, 10, 10, 14, 15, 15, 18, 19, 21, 22, 22, 23, 25
Grupo 3 (13)	8, 8, 8, 8, 8, 8, 9, 10, 10, 10, 11, 17, 19

Este exemplo, caracterizado pela censura do tipo I, é utilizado no Capítulo 2 para ilustrar as técnicas não-paramétricas para dados de sobre-

vivência.

1.5.3 Dados de Leucemia Pediátrica

Esses dados foram obtidos a partir de um estudo de crianças com leucemia, desenvolvido pelo Grupo Cooperativo Mineiro para Tratamento de Leucemias Agudas. Este é um estudo descritivo. A leucemia aguda é a neoplasia de maior incidência na população com menos de 15 anos de idade. Calcula-se que, nesta faixa etária, a incidência anual gira em torno de 5 a 6 casos novos por 100 mil crianças, sendo a grande maioria dos casos de Leucemia Linfoblástica Aguda (LLA).

Apesar do progresso alcançado no tratamento, em particular, da leucemia linfoblástica, as leucemias agudas continuam sendo a causa mais comum de morte por neoplasia. O objetivo do tratamento médico de uma criança com LLA é obter longos períodos de sobrevida livre da doença, o que, muitas vezes, significa sua *cura*. Os avanços terapêuticos obtidos nos últimos 25 anos têm sido grandes na LLA. Na década de 60, menos de 1% das crianças com LLA sobreviviam mais de 5 anos após o diagnóstico. Atualmente, com a intensificação da quimioterapia para os grupos com prognóstico mais desfavorável, 60 a 70% do total de crianças com diagnóstico de LLA são sobreviventes de longo prazo e encontram-se provavelmente *curadas*. Nos grupos de melhor prognóstico, as proporções de *cura* já se situam no patamar de 90%.

Com o objetivo de entender melhor quais fatores afetam o tempo de sobrevivência de uma criança brasileira com LLA, um grupo de 128 crianças, com idade inferior a 15 anos, foi acompanhado no período de 1988 a 1992, em alguns hospitais de Belo Horizonte. A variável resposta de interesse foi o tempo a partir da remissão (ausência da doença) até a recidiva ou morte (a que ocorrer primeiro). Das 128 crianças, 120 entraram em remissão, e são estas que formam o conjunto de dados em estudo. Os fatores registrados para cada criança e que compõem o banco de dados são os seguintes:

idade, peso, estatura, contagem de leucócitos, porcentagem de linfoblastos, porcentagem de vacúolos, fator de risco e indicador de sucesso da remissão.

No Capítulo 5, os dados desse estudo são analisados por meio do modelo de Cox. No Capítulo 6, estes são utilizados para ilustrar o modelo de Cox estratificado. Informações adicionais sobre o estudo podem ser encontradas em Viana et al. (1994) e Colosimo et al. (1992).

1.5.4 Dados de Sinusite em Pacientes Infectados pelo HIV

O estudo da epidemia da AIDS é uma área de intensa pesquisa e vários trabalhos já estão registrados na literatura. A maioria deles tem como foco principal de atenção o tempo de vida de pacientes. Um estudo desenvolvido pela Profa. Denise Gonçalves do Departamento de Otorrinolaringologia da UFMG teve como interesse a ocorrência de manifestações otorrinolaringológicas em pacientes HIV positivos. O objetivo mais específico, e que é explorado neste texto, é verificar a hipótese de que a infecção pelo HIV aumenta o risco de ocorrência de sinusite.

Nesse estudo foram utilizadas informações provenientes de 91 pacientes HIV positivo e 21 HIV negativo, somando assim 112 pacientes estudados. Estes pacientes foram acompanhados no período compreendido entre março de 1993 e fevereiro de 1995. A classificação do paciente quanto à infecção pelo HIV seguiu os critérios do CDC (*Centers of Disease Control*, 1987). Os pacientes foram classificados como: HIV soronegativo (não possuem o HIV), HIV soropositivo assintomático (possuem o vírus mas não desenvolveram o quadro clínico de AIDS), com ARC, *AIDS Related Complex* (apresentam baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS), ou com AIDS (já desenvolveram infecções oportunistas que definem AIDS, segundo os critérios do CDC de 1987). Esta é a principal covariável a ser considerada no estudo. Ela é dependente do tempo, pois os pacientes mudam de classificação ao longo do estudo. Esta característica requer técnicas especializadas que são apresentadas no

Capítulo 6. Outras covariáveis neste estudo, como contagem de células CD4 e CD8, também são dependentes do tempo. No entanto, elas foram somente medidas no início do estudo e, ainda, ocorreu a falta de registro de ambas as contagens para em torno de 37% dos pacientes. Desse modo, elas não foram incluídas nas análises.

A cada consulta, a classificação do paciente foi reavaliada. Cada paciente foi acompanhado através de consultas trimestrais. A frequência mediana de consultas foi 4. A resposta de interesse foi o tempo, em dias, contado a partir da primeira consulta até a ocorrência de sinusite. O objetivo foi identificar fatores de risco para esta manifestação. Os possíveis fatores de risco incluídos no estudo estão listados na Tabela 1.3.

Tabela 1.3: Covariáveis medidas no estudo de ocorrência de sinusite.

Idade do Paciente	Foi medida em anos
Sexo do Paciente	0 - Masculino 1 - Feminino
Grupos de Risco	1 - Paciente HIV Soronegativo 2 - Paciente HIV Soropositivo Assintomático 3 - Paciente com ARC 4 - Paciente com AIDS
Atividade Sexual	1 - Homossexual 2 - Bissexual 3 - Heterossexual
Uso de Droga	1 - Sim
Injetável	2 - Não
Uso de Cocaína	1 - Sim
por Aspiração	2 - Não

Foram registrados 23 valores perdidos para as covariáveis Atividade Sexual, Uso de Droga e Uso de Cocaína. Mais informações sobre este

estudo podem ser encontradas em Gonçalves (1995) e Colosimo e Vieira (1996).

1.5.5 Dados de Aleitamento Materno

As Organizações Internacionais de Saúde recomendam o leite materno como a única fonte de alimentação para crianças entre 4 e 6 meses de idade. Identificar fatores determinantes do aleitamento materno em diferentes populações é, portanto, fundamental para alcançar tal recomendação.

Os professores Eugênio Goulart e Cláudia Lindgren do Departamento de Pediatria da UFMG realizaram um estudo no Centro de Saúde São Marcos, localizado em Belo Horizonte, com o objetivo principal de conhecer a prática do aleitamento materno de mães que utilizam este centro, assim como os possíveis fatores de risco ou de proteção para o desmame precoce. Um inquérito epidemiológico composto por questões demográficas e comportamentais foi aplicado a 150 mães de crianças menores de 2 anos de idade. A variável resposta de interesse foi estabelecida como sendo o tempo máximo de aleitamento materno, ou seja, o tempo contado a partir do nascimento até o desmame completo da criança.

Uma análise estatística utilizando modelos paramétricos e semiparamétricos é realizada nos Capítulos 4 e 5 para estes dados. Desta forma, pode-se comparar os resultados obtidos usando-se ambos os modelos.

1.5.6 Dados Experimentais utilizando Camundongos

Um estudo laboratorial foi realizado para investigar o efeito protetor do fungo *Saccharomyces boulardii* em ratos debilitados imunologicamente. O estudo utilizou 93 ratos provenientes do mesmo biotério. Inicialmente, o sistema imunológico dos ratos foi debilitado quimicamente e, a seguir, 4 tratamentos (controle e o fungo nas dosagens: 10mg, 1mg e 0,1mg) foram alocados aleatoriamente a cada animal. Como resposta de interesse foi considerado o tempo de vida, medido em dias, após a aplicação do tratamento.

O estudo teve por objetivo comparar os tratamentos controlando pelo peso inicial do rato. Uma característica desses dados é a presença de empates. Existem 61 tempos de censura e 13 tempos de falha distintos entre as 32 mortes observadas durante o período do estudo. A possibilidade de ajustar um modelo de regressão discreto para um conjunto de dados com vários empates é discutida no Capítulo 8.

1.5.7 Dados de Câncer de Mama

Com o objetivo de pesquisar duas terapias: (a) somente radioterapia e (b) radioterapia em conjunto com quimioterapia, um estudo retrospectivo foi realizado com 94 mulheres com diagnóstico precoce de câncer de mama. Um total de 46 delas recebeu a primeira terapia e as demais receberam a segunda. As pacientes foram acompanhadas a cada 4-6 meses e, em cada visita, foram registrados: a ocorrência da retração da mama (nenhuma, moderada ou severa) e o tempo até o aparecimento de uma retração moderada ou severa da mama. Como as visitas foram realizadas em alguns tempos aleatórios, não se sabe com exatidão quando a primeira retração da mama ocorreu; sabe-se somente que esta ocorreu entre duas das visitas realizadas. Por outro lado, o que se sabe a respeito das pacientes que não apresentaram retração da mama até a última visita é que o evento não ocorreu até aquele momento e que, caso venha a ocorrer, será a partir daquele momento em diante. Este exemplo é analisado no Capítulo 8, em que é abordada a análise de dados de sobrevivência intervalar. Informações adicionais sobre este estudo podem ser encontradas em Klein e Moeschberger (2003).

1.5.8 Dados de Tempo de Vida de Mangueiras

No período de 1972 a 1992, um ensaio em delineamento em blocos ao acaso foi conduzido no Departamento de Horticultura da ESALQ/USP. O objetivo foi verificar a resistência das mangueiras a uma praga denominada seca da mangueira, que mata a planta. O interesse concreto era identificar novas

mangueiras obtidas a partir de enxertos, resistentes à seca da mangueira. Um experimento fatorial completamente aleatorizado foi realizado com 6 copas enxertadas sobre 7 porta-enxertos (fatorial 6×7). Todas as 42 combinações foram replicadas em 5 blocos diferentes, totalizando 210 unidades experimentais. O estudo teve início em 1971 e a resposta de interesse foi o tempo de vida das mangueiras. O experimento foi visitado 12 vezes durante o período do estudo e foi registrada a condição de cada unidade experimental (viva ou morta). Os dados provenientes desse estudo são de natureza intervalar, ou seja, o evento de interesse (morte da mangueira) acontece entre duas visitas consecutivas e o tempo exato da morte é desconhecido. Este exemplo é analisado no Capítulo 8, que é dedicado a dados de sobrevivência intervalar e grupados. Mais informações sobre este estudo podem ser encontradas em Chalita et al. (1999) e Giolo et al. (2009).

1.6 Especificando o Tempo de Sobrevivência

A variável aleatória não-negativa T , usualmente contínua, que representa o tempo de falha, é geralmente especificada em análise de sobrevivência pela sua função de sobrevivência ou pela função de taxa de falha. Estas duas funções, e funções relacionadas, que são extensivamente usadas na análise de dados de sobrevivência são apresentadas a seguir.

1.6.1 Função de Sobrevivência

Esta é uma das principais funções probabilísticas usadas para descrever estudos de sobrevivência. A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver ao tempo t . Em termos probabilísticos, isto é escrito como:

$$S(t) = P(T \geq t).$$

Em conseqüência, a função de distribuição acumulada é definida como a probabilidade de uma observação não sobreviver ao tempo t , isto é, $F(t) = 1 - S(t)$.

Na Figura 1.2 pode ser observada a forma típica de duas funções de sobrevivência. Estas curvas que, nesse caso, representam as funções de sobrevivência de dois grupos de pacientes, o grupo 1, tratado com a droga A e o grupo 2, com a droga B, fornecem informações importantes. Note, por exemplo, que o tempo de vida dos pacientes do grupo 1 é superior ao dos pacientes do grupo 2 na maior parte do tempo de acompanhamento. Para os pacientes do grupo 1, o tempo para o qual cerca de 50% (tempo mediano) deles morrem é de 20 anos, enquanto que, para os pacientes do grupo 2, este tempo é menor (10 anos). Outra informação importante e possível de ser retirada desta figura é o percentual de pacientes que ainda estão vivos até um determinado tempo de interesse. Por exemplo, para os pacientes do grupo 1, cerca de 90% deles ainda estão vivos após 10 anos do início do estudo, enquanto que, para os do grupo 2, apenas 50%.

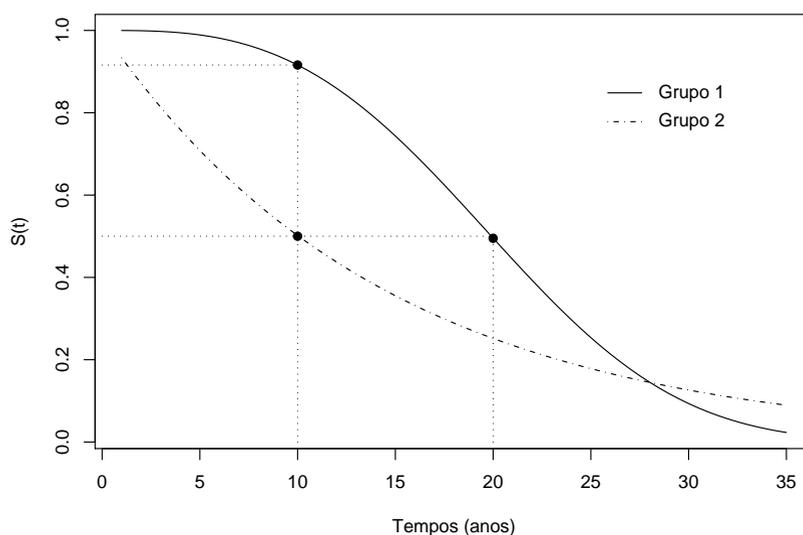


Figura 1.2: Funções de sobrevivência para dois grupos de pacientes.

1.6.2 Função de Taxa de Falha

A probabilidade da falha ocorrer em um intervalo de tempo $[t_1, t_2)$ pode ser expressa em termos da função de sobrevivência como:

$$S(t_1) - S(t_2).$$

A taxa de falha no intervalo $[t_1, t_2)$ é definida como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de t_1 , dividida pelo comprimento do intervalo. Assim, a taxa de falha no intervalo $[t_1, t_2)$ é expressa por:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1) S(t_1)}. \quad (1.1)$$

De forma geral, redefinindo o intervalo como $[t, t + \Delta t)$, a expressão (1.1) assume a seguinte forma:

$$\lambda(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}.$$

Assumindo Δt bem pequeno, $\lambda(t)$ representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t . Observe que as taxas de falha são números positivos, mas sem limite superior. A função de taxa de falha $\lambda(t)$ é bastante útil para descrever a distribuição do tempo de vida de pacientes. Ela descreve a forma em que a taxa instantânea de falha muda com o tempo.

A função de taxa de falha de T é, então, definida como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (1.2)$$

A Figura 1.3 mostra três funções de taxa de falha. A função crescente indica que a taxa de falha do paciente aumenta com o transcorrer do tempo. Este comportamento mostra um efeito gradual de envelhecimento. A função constante indica que a taxa de falha não se altera com o passar do tempo. A função decrescente mostra que a taxa de falha diminui à medida que o tempo passa.

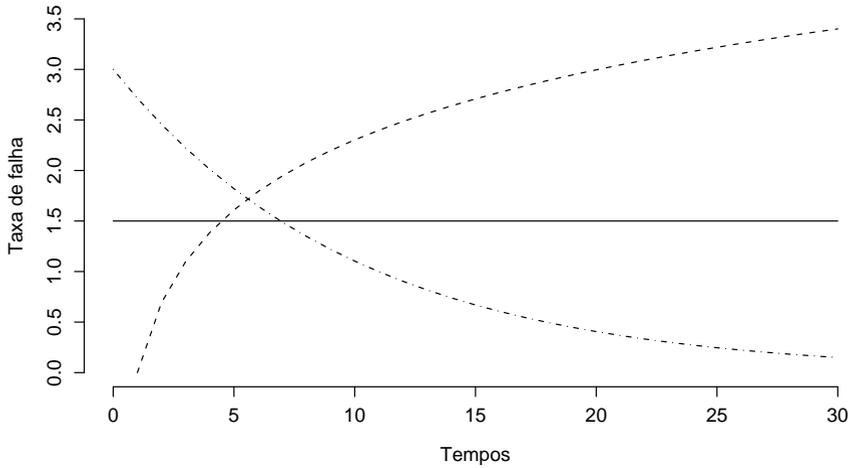


Figura 1.3: Funções de taxa de falha -- crescente, — constante e ---- decrescente.

Sabe-se, ainda, que a taxa de falha para o tempo de vida de seres humanos é uma combinação das curvas apresentadas na Figura 1.3 em diferentes períodos de tempo. Ela é conhecida como *curva da banheira* e tem uma taxa de falha decrescente no período inicial, representando a mortalidade infantil, constante na faixa intermediária e crescente na porção final. Uma representação desta curva é mostrada na Figura 1.4.

A função de taxa de falha é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções de taxa de falha podem diferir drasticamente. Desta forma, a modelagem da função de taxa de falha é um importante método para dados de sobrevivência.

1.6.3 Função de Taxa de Falha Acumulada

Outra função útil em análise de dados de sobrevivência é a função de taxa de falha acumulada. Esta função, como o próprio nome sugere, fornece a taxa de falha acumulada do indivíduo e é definida por:

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

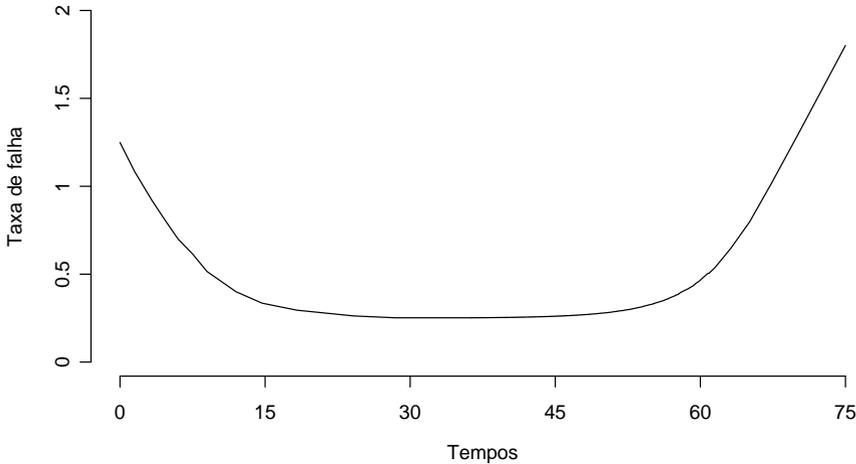


Figura 1.4: Representação da função de taxa de falha conhecida por *curva da banheira*.

A função de taxa de falha acumulada, $\Lambda(t)$, não tem uma interpretação direta, mas pode ser útil na avaliação da função de maior interesse que é a de taxa de falha, $\lambda(t)$. Isto acontece essencialmente na estimação não-paramétrica em que $\Lambda(t)$ apresenta um estimador com propriedades ótimas e $\lambda(t)$ é difícil de ser estimada.

1.6.4 Tempo Médio e Vida Média Residual

Outras duas quantidades de interesse em análise de sobrevivência são: o tempo médio de vida e a vida média residual. A primeira é obtida pela área sob a função de sobrevivência. Isto é,

$$t_m = \int_0^{\infty} S(t) dt.$$

Já a vida média residual é definida condicional a um certo tempo de vida t . Ou seja, para indivíduos com idade t esta quantidade mede o tempo médio restante de vida e é, então, a área sob a curva de sobrevivência à direita do tempo t dividida por $S(t)$. Isto é,

$$\text{vmr}(t) = \frac{\int_t^\infty (u-t)f(u)du}{S(t)} = \frac{\int_t^\infty S(u)du}{S(t)}$$

sendo $f(\cdot)$ a função de densidade de T . Observe que $\text{vmr}(0) = t_m$.

1.6.5 Relações entre as Funções

Para T uma variável aleatória contínua e não-negativa, tem-se, em termos das funções definidas anteriormente, algumas relações matemáticas importantes entre elas, a saber:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t)),$$

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log S(t)$$

e

$$S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u) du\right\}.$$

Tais relações mostram que o conhecimento de uma das funções, por exemplo $S(t)$, implica no conhecimento das demais, isto é, de $F(t)$, $f(t)$, $\lambda(t)$ e $\Lambda(t)$.

Outras relações envolvendo estas funções são as seguintes:

$$S(t) = \frac{\text{vmr}(0)}{\text{vmr}(t)} \exp\left\{-\int_0^t \frac{du}{\text{vmr}(u)}\right\}$$

e

$$\lambda(t) = \left(\frac{d \text{vmr}(t)}{dt} + 1\right) / \text{vmr}(t).$$

1.7 Exercícios

1. Suponha que seis ratos foram expostos a um material cancerígeno. Os tempos até o desenvolvimento do tumor de um determinado tamanho são registrados para os ratos. Os ratos A, B e C desenvolveram os tumores em 10, 15 e 25 semanas, respectivamente. O rato D morreu acidentalmente sem tumor na vigésima semana de observação. O estudo terminou com 30 semanas sem os ratos E e F apresentarem tumor.
 - (a) Defina cuidadosamente a resposta do estudo.
 - (b) Identifique o tipo de resposta (falha ou censura) observado para cada um dos ratos no estudo.

2. Um número grande de indivíduos foi acompanhado para estudar o aparecimento de um certo sintoma. Os indivíduos foram incluídos ao longo do estudo e foi considerada como resposta de interesse a idade em que este sintoma apareceu pela primeira vez. Para os seis indivíduos selecionados e descritos a seguir, identifique o tipo de censura apresentado.
 - (a) O primeiro indivíduo entrou no estudo com 25 anos já apresentando o sintoma.
 - (b) Outros dois indivíduos entraram no estudo com 20 e 28 anos e não apresentaram o sintoma até o encerramento do estudo.
 - (c) Outros dois indivíduos entraram com 35 e 40 anos e apresentaram o sintoma no segundo e no sexto exames, respectivamente, após terem entrado no estudo. Os exames foram realizados a cada dois anos.
 - (d) O último indivíduo selecionado entrou no estudo com 36 anos e mudou da cidade depois de 4 anos sem ter apresentado o sintoma.

3. Mostre que $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t))$.
4. Mostre que $\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t)$.
5. Mostre que $\text{vmr}(t) = \frac{\int_t^\infty (u-t)f(u)du}{S(t)} = \frac{\int_t^\infty S(u)du}{S(t)}$.
(Sugestão: utilize uma integral por partes sabendo que $f(u)du = -\frac{d}{du}S(u)$).
6. Suponha que a taxa de falha da variável aleatória tempo de falha T seja expressa pela função linear $\lambda(t) = \beta_0 + \beta_1 t$, com β_0 e $\beta_1 > 0$. Obtenha $S(t)$ e $f(t)$.
7. Suponha que a vida média residual de T seja dada por $\text{vmr}(t) = t+10$. Obtenha $E(T)$, $\lambda(t)$ e $S(t)$.
8. Em cada um dos exemplos descritos na Seção 1.5, identifique o tempo inicial, a escala de medida e o evento de interesse.

ANÁLISE DE SOBREVIVÊNCIA APLICADA

Enrico Antônio Colosimo

Professor do Departamento de Estatística da UFMG. Ph.D. em Estatística pela University of Wisconsin-Madison. Editor associado da *Revista Brasileira de Estatística*. Coautor do livro *Confiabilidade: análise de tempo de falha e testes de vida acelerados*. As suas áreas de interesse são métodos estatísticos em análise de sobrevivência, confiabilidade/manutenção e dados longitudinais.

Suely Ruiz Giolo

Professora do Departamento de Estatística da UFPR. Licenciada em Matemática e graduada em Estatística pela Unesp, mestre em Estatística pela Unicamp e doutora em Estatística e Experimentação Agronômica pela ESALQ/USP e pela Lancaster University, Inglaterra. Entre suas áreas de interesse estão análise de sobrevivência, de dados longitudinais e de dados categóricos.

Sobre o livro

Neste livro são apresentados conceitos básicos, técnicas não paramétricas e os principais modelos probabilísticos e de regressão utilizados na análise de dados de sobrevivência. Dentre os modelos, podem ser citados o de Cox, o aditivo de Aalen, o de fragilidade gama e alguns modelos para dados de censura intervalar e grupados. Diversos exemplos reais provenientes, em sua maioria, de situações clínicas são analisados e ilustram as técnicas e modelos apresentados. Para obtenção dos resultados estatísticos é utilizado o pacote estatístico R, cujos comandos são descritos no texto. De modo geral, este livro pode ser utilizado por alunos de graduação e de pós-graduação em Estatística, bem como por alunos, profissionais e pesquisadores de outras áreas (medicina, biologia etc.) que tenham interesse em análise de sobrevivência.

www.blucher.com.br

ISBN 978-85-212-0384-1

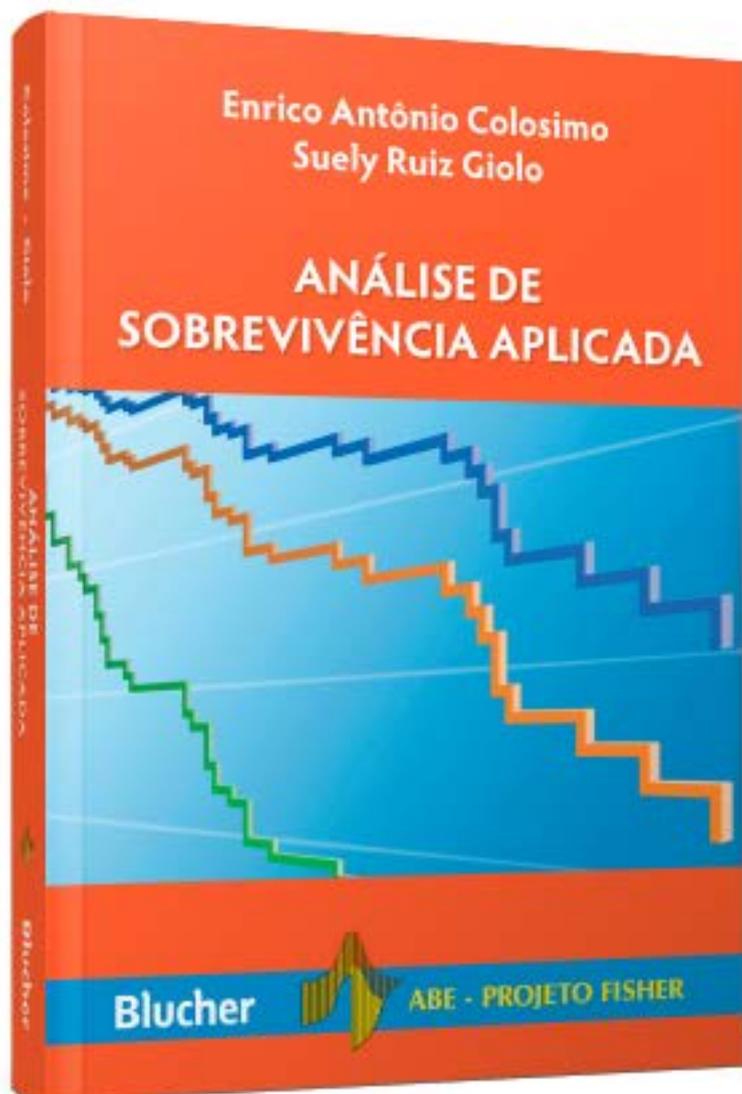


9 788521 203841



ABE - PROJETO FISHER

Blucher



Clique aqui e:

[VEJA NA LOJA](#)

Análise de Sobrevivência Aplicada

Enrico Antônio Colosimo , Suely Ruiz Giolo

ISBN: 9788521203841

Páginas: 392

Formato: 17x24 cm

Ano de Publicação: 2006

Peso: 0.625 kg
