

TÉCNICAS DE ANÁLISE DE SOBREVIVÊNCIA, COM APLICAÇÕES. Elton Rafael Maurício da Silva Pereira, José Silvio Govone. – Inter-Áreas – Probabilidade e Estatística - Departamento de Estatística, Matemática Aplicada e Computação - Licenciatura Plena em Matemática - IGCE/UNESP – Campus de Rio Claro.

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas duas décadas, devido ao desenvolvimento e aprimoramento de técnicas estatísticas, combinadas com computadores cada vez mais velozes. A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta, o termo “análise de sobrevivência” refere-se basicamente a situações médicas envolvendo dados censurados. Entretanto, podemos encontrar a presença de situações similares em engenharia e em ciências sociais.

Os dados de sobrevivência são caracterizados pelo tempo de falha o qual é formado pelos seguintes elementos: tempo inicial, escala de medida e o evento de interesse, e frequentemente pelas censuras, que juntos constituem a resposta. O tempo inicial deve ser precisamente definido, a escala de medida é quase sempre o tempo real ou “de relógio”; contudo podemos recorrer a outras, tais como, na engenharia encontramos como escalas de medida: número de ciclos, quilometragem de um carro ou qualquer outra medida de carga.

E o terceiro elemento é o evento de interesse o qual é em geral indesejável e comumente chamado de falha.

Em estudos de sobrevivência é muito importante definirmos de forma clara e precisa o que vem a ser a falha. 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”. É necessário definirmos claramente este evento antes do início do estudo, para melhor compreensão poderíamos dizer que o a falha ocorrerá a partir do momento em que o produto atingir mais do que uma determinada concentração de microorganismos por mm^2 de área do produto.

Os objetivos de uma análise de sobrevivência envolvendo dados de sobrevivência estão geralmente relacionados, em medicina, à identificação de fatores de prognóstico para uma certa doença ou à comparação de tratamentos em estudo clínico enquanto controlado por outros fatores.

Os estudos clínicos de sobrevivência geralmente terminam antes que todos os indivíduos no estudo venham a falhar. Uma característica que se observa no estudo é a observação parcial ou incompleta, tais observações são chamadas de censuras. Temos alguns mecanismos de censura que se destacam em estudos clínicos, tais como: Censura do Tipo I é aquela em que o estudo será terminado após um período pré-estabelecido de tempo. Censura Tipo II é aquela em que o estudo será terminado após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos. E um terceiro mecanismo ocorre quando um paciente é retirado do estudo sem ter ocorrido a falha.

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 $\delta_i = 1$, se t_i é um tempo de falha ou $\delta_i = 0$, se t_i é um tempo censurado.

Função de Sobrevivência

Esta é uma das principais funções probabilísticas usadas para descrever estudos de sobrevivência. Ela 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 . Podemos descrevê-la como $S(t) = P(T \geq t)$.

Função Taxa de Falha ou de Risco

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. Podemos expressá-la da seguinte forma: $S(t_1) - S(t_2) / (t_2 - t_1) S(t_1)$.

Por mais complexo que seja o estudo, as respostas às perguntas de interesse são dadas a partir de um conjunto de dados de sobrevivência os quais são analisados a partir de algumas técnicas, dentre elas, algumas são objeto de estudo do presente trabalho, tais como: Estimador Kaplan – Meier, Estimador de Nelson-Aalen, Estimador da Tabela de Vida ou Atuarial.

A utilização da curva de Kaplan-Meier nos informa a probabilidade estimada de sobrevivência por um determinado tempo. O estimador de Kaplan Méier é uma adaptação da função de sobrevivência empírica que, na ausência de censuras é definida como:

$S'(t) = (\text{número de observações que não falharam até o tempo } t) / (\text{número total de observações no estudo})$

A seguir apresentamos um exemplo da utilização da técnica de Kaplan-Meier.

Dados de Hepatite (Colosimo, Enrico Antônio, 2006)

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 abaixo. O símbolo + indica censura. Este exemplo é caracterizado pela censura do tipo aleatória.

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

O procedimento para se obter a estimativa de Kaplan-Meier envolve uma seqüência de passos, em que o próximo depende do anterior. Isto significa, por exemplo que: $S(5)=P(T \leq 5)=P(T \leq 1, T \leq 5)=P(T \leq 1)P(T \leq 5 | T \leq 1)$.

Desta forma, para o indivíduo sobreviver por 5 semanas, ele vai precisar sobreviver, em um primeiro passo, à primeira semana e depois sobreviver à quinta semana, sabendo-se que ele sobreviveu à primeira semana. Os tempos de 1 e 5 semanas foram tomados por serem os dois primeiros tempos distintos de falha nos dados do grupo esteróide. Os passos são gerados a partir de intervalos definidos pela ordenação dos tempos de falha de forma que cada um deles começa em um tempo observado e termina no próximo tempo. A tabela abaixo apresenta os tempos ordenados e mostra a existência de 6 intervalos, iniciando com $[0,1)$, até o sexto intervalo que é $[10,16)$. O limite superior deste último intervalo é definido como sendo 16 semanas por ser este o maior tempo de acompanhamento do estudo.

t_j	Intervalos	d_j	n_j	$S'(t_j +)$
0	$[0,1)$	0	14	1,000
1	$[1,5)$	3	14	0,786
5	$[5,7)$	1	9	0,698
7	$[7,8)$	1	8	0,611
8	$[8,10)$	1	7	0,524
10	$[10,16)$	1	6	0,437

Todos os indivíduos estavam vivos em $t=0$ e se mantêm até a primeira morte que ocorre em $t=1$ semana. Então, a estimativa de $S(t)$ deve ser 1 neste intervalo compreendido entre 0 e 1 semana. No valor correspondente a 1 semana, a estimativa deve ter caído devido a três mortes que ocorrem neste tempo. No segundo intervalo, $[1,5)$, existem, então, 14 indivíduos que estavam vivos (sob risco) antes de $t=1$ e 3 morrem. Desta forma, a estimativa da probabilidade condicional de morte neste intervalo é $3/14$ e a probabilidade de sobreviver é $1-3/14$. Isto pode ser escrito como:

$S'(1)=P(T \leq 0)P(T \leq 1 | T \leq 0)=(1)(11/14)=0,786$.

O estimador de Kaplan-Meier é definido como:

$$\prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right),$$

d_j o número de falhas em t_j , $j=1, \dots, k$, e n_j o número de indivíduos sob risco em t_j , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j .

A grande diferença entre os estimadores está no número de intervalos utilizados para a construção de cada um deles. O estimador Kaplan-Meier e o de Nelson-Aalen são sempre baseados em um número de intervalos igual ao número de tempos de falha distintos, enquanto que, na tabela de vida, os tempos de falha são agrupados em intervalos, de forma arbitrária.

Aplicações das técnicas estudadas estão sendo feitas, utilizando dados da cidade de Rio Claro, comparando curvas de sobrevivência referentes às idades em que ocorrem eventos de mortes violentas (homicídios, suicídios e acidentes) e idades onde ocorrem mortes naturais. Tais dados foram fornecidos pelo Cartório do Registro Civil, de Rio Claro.

Referências Bibliográficas

Colosimo, Enrico Antônio e Giolo, Suely Ruiz (2006), *Análise de Sobrevivência*. Editora Edgard Blücher LTDA, São Paulo.
