

Modelagem de tempo de sobrevida via método bootstrap

C. M. Silva^{1,2}; R. S. Amaral²; J. W. Vieira^{2,3}; A. N.C. Silva²; J. A. S. Júnior²; E. S. Alcoforado²

¹*Departamento de Matemática, Universidade de Pernambuco-Campus Petrolina, 56328-586, Petrolina-Pe, Brasil*

²*Departamento de Energia Nuclear, Universidade Federal de Pernambuco, 50740-540, Recife-Pe, Brasil*

³*Instituto Federal de Educação, Ciência e Tecnologia, 50740-540, Recife-Pe, Brasil*

cleomacio@ig.com.br

(Recebido em 02 de setembro de 2011; aceito em 04 de outubro de 2011)

Em estudos estatísticos de análise de sobrevida, o tempo não é normalmente distribuído, pois é uma variável contínua que não admite valores negativos. Sendo assim, as funções densidade de probabilidade exponencial, Weibull e lognormal são as mais utilizadas para modelar tempo de sobrevida. Cada uma dessas funções possui suas limitações e complexidades quando utilizadas para estimar tempo de sobrevida. O método bootstrap consiste de um procedimento estatístico computacionalmente intensivo que permite avaliar diversas estatísticas, com base nos dados obtidos da amostra. Portanto, dentro desse contexto, o objetivo do presente estudo foi apresentar um método alternativo para modelar tempo de sobrevida utilizando o método bootstrap. Para tanto, foi desenvolvido um programa computacional na linguagem C++. Os resultados obtidos indicaram que o uso do método bootstrap na modelagem de tempo de sobrevida, evitou as interferências estatísticas resultantes da escolha de um modelo de distribuição de probabilidade.

Palavras-chave: computação científica, estatística computacional, modelos matemáticos.

In studies of statistical analysis of survival time is not normally distributed, it is a continuous variable that does not allow negative values. Thus, the probability density functions exponential, Weibull and lognormal are the most commonly used to model survival time. Each of these functions has its limitations and complexities when used to estimate median survival time. The bootstrap method consists of a computationally intensive statistical procedure that allows evaluating several statistics based on data obtained from the sample. Therefore, in this context, the objective of this study was to present an alternative method to model the average time in the statistical analysis of survival using the bootstrap method. To this end, we developed a computer program in C++. The results indicated that the use of the bootstrap method in modeling the median survival time, avoided the interference statistics resulting from the choice of a probability distribution model.

Keywords: scientific computing, statistical computing, mathematical models.

1. INTRODUÇÃO

Em todo procedimento estatístico de análise de sobrevida, o tempo é uma variável muito importante. A análise de sobrevida, também chamada de análise de sobrevivência, é utilizada quando o tempo for o objeto de interesse, seja este interpretado como o tempo até a ocorrência de um evento ou o risco de ocorrência de um evento por unidade de tempo [1]. Dentro do contexto epidemiológico existem três dimensões distintas de tempo. A primeira diz respeito à idade do indivíduo. Predições realizadas em bases individuais sobre a evolução da grande maioria das condições de morbidade, geralmente, têm na idade uma variável importante para o risco da ocorrência do desfecho [1]. A segunda dimensão contabiliza a duração em uma determinada condição. Esta é a dimensão natural a ser utilizada em ensaios clínicos. A terceira dimensão refere-se ao tempo calendário propriamente dito. O acesso à descrição da taxa de incidência como uma função do tempo calendário, permite o estudo da evolução temporal de padrões de morbidade e a descrição de seus determinantes fornecendo as bases empíricas das ações em saúde pública [1].

Uma característica especial da variável aleatória tempo de sobrevida é que esta é uma quantidade contínua não-negativa. Assim, a distribuição normal, que permite valores negativos, não é adequada para modelar o tempo de sobrevida [1]. Além disso, frequentemente, o tempo de sobrevida apresenta forte assimetria, com uma grande cauda à esquerda. Esta forma ocorre por que grande parte dos tempos de sobrevida tem valores pequenos com poucos indivíduos com tempos muito longos, gerando assim a assimetria [1]. Ao considerar a possibilidade de se adotar uma distribuição teórica para o tempo de sobrevida, as funções de densidade de probabilidade mais utilizada, por se adaptarem a uma grande variedade de situações, são: exponencial, Weibull e lognormal [1].

Historicamente, a distribuição exponencial foi a mais utilizada para modelar tempo de sobrevida. Apesar da simplicidade matemática do modelo exponencial, a suposição de risco constante no tempo é muito pouco plausível na maioria dos fenômenos da saúde. Em algumas situações particulares, porém, o modelo exponencial pode ser uma aproximação válida: quando o tempo de acompanhamento é curto o suficiente para que o risco naquele período possa ser considerado constante [1]. Para a maioria dos fenômenos na área da saúde, é mais correto considerar que o risco não varia linearmente com o tempo. Assim, a função de Weibull, atualmente, é a mais utilizada para modelar tempo de sobrevida. Esta função é uma generalização da função exponencial. Porém, a função de Weibull oferece dificuldades matemáticas na determinação do valor do tempo médio. Outra possibilidade em análise de sobrevida considera que o tempo possui distribuição lognormal, isto é, que o logaritmo do tempo de sobrevida tem distribuição normal. Uma propriedade pouco atrativa da distribuição lognormal refere-se ao comportamento decrescente da função de risco para valores grandes do tempo de sobrevida. Esse decréscimo é pouco plausível na maioria das situações na área de saúde [1].

As funções de densidade de probabilidade exponencial, Weibull e lognormal possuem suas facilidades, limitações e complexidades na modelagem de tempo de sobrevida. Apesar das aplicações dessas funções em estudos estatísticos de análise de sobrevida, é necessário testar novos procedimentos estatísticos para modelar a variável tempo, de tal maneira que haja confiabilidade e credibilidade nos resultados obtidos. Nos últimos anos, o método bootstrap vem sendo aplicado com grandes êxitos em diferentes áreas do conhecimento humano. Sendo assim, este método pode constituir-se numa ferramenta estatística importante para modelar tempo de sobrevida.

O termo *bootstrap* surgiu da frase *to pull oneself up one's bootstrap* retirada de *Adventures of Baron Munchausen* by Rodolph Erich Raspe, XVII century: *The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought of picking himself up by his own bootstraps* [2]. Esse texto relata uma situação em que o Barão Munchausen está afundando em um lago e vendo que tudo estava perdido, pensa que conseguirá emergir puxando os cadarços dos próprios sapatos. O sentido estatístico do termo é passar a ideia de que, em situações difíceis, devem-se tentar as mais variadas soluções possíveis a partir dos dados originais. Em estatística, situações difíceis podem ser vistas como os problemas de soluções analíticas complexas. As variadas soluções possíveis seria a utilização de uma metodologia com grande quantidade de cálculos, objetivando extrapolar os resultados a partir de um pequeno conjunto de dados. Com o uso sistematizado de ferramentas computacionais, a solução para esses casos, é obtida substituindo-se a resolução analítica pelo poder de processamento dos computadores através do método de amostragem bootstrap.

Na análise de sobrevida, a dispersão dos valores do tempo, geralmente, é bastante elevada, fazendo com que se prefira utilizar mais o tempo mediano do que o tempo médio. Entretanto, caso a variação seja grande, a utilização da média aritmética, como valor mais representativo, seria totalmente inadequada, devido à discrepância entre os valores. Em estudos de saúde não existem procedimentos estatísticos utilizados para reduzir os efeitos causados pelos valores anômalos sobre a média aritmética. Devido às flutuações estatísticas causadas pelos valores *outliers* (anômalos), utilizam-se, geralmente, a mediana como valor mais representativo do conjunto de dados obtidos da amostra de tempo de sobrevida, pois, ela não é afetada pelos valores discrepantes. Dois fatores importantes limitam a utilização da mediana em análise de sobrevida: (1) a mediana é sempre menor do a média aritmética e (2) a mediana discrimina os

valores extremos. Tais limitações fazem com que informações importantes sejam perdidas durante a análise estatística dos dados. Sendo assim, a mediana não converge adequadamente para o valor mais representativo do conjunto de dados de tempo de sobrevivência. Por outro lado, o método de reamostragem bootstrap, quando aplicado na reamostragem dos dados originais obtidos da amostra, fornece uma média aritmética resistente às flutuações causadas pelos efeitos dos valores anômalos. Neste caso, a reamostragem é utilizada para diminuir a assimetria, acomodando os valores de tal maneira, que a discrepância em torno da média aritmética passa a ser a menor possível [2]. Portanto, o objetivo da presente pesquisa foi desenvolver um programa computacional na linguagem C++ com a finalidade de estimar tempo de sobrevivência.

2. MATERIAIS E MÉTODOS

As funções densidade de probabilidade exponencial, lognormal e Weibull foram utilizadas no presente estudo com a finalidade de comparar os resultados obtidos do tempo de análise de sobrevivência com aquele calculado pelo método bootstrap.

2.1. A distribuição exponencial

A distribuição exponencial é um tipo de distribuição contínua de probabilidade, representada por um parâmetro λ . Historicamente, este tipo de função foi a mais amplamente utilizada na modelagem de tempo de sobrevida [1]. Ao assumir que a variável tempo de sobrevida possui uma distribuição exponencial, a função densidade de probabilidade é calculada pela Equação 1 [1].

$$f(t, \lambda) = \lambda \exp(-\lambda t), t > 0 \quad (1)$$

No modelo exponencial o tempo médio e o tempo mediano são calculados, respectivamente pelas Equações 2 e 3 [1].

$$t_{\text{médio}} = \frac{1}{\lambda} \quad (2)$$

$$t_{\text{mediano}} = \frac{\ln 2}{\lambda} \quad (3)$$

2.2. A distribuição lognormal

A variável t (tempo de sobrevivência), com valores positivos, tem distribuição lognormal com parâmetros μ e σ^2 , com $\sigma^2 > 0$, se $y = \ln(t)$ tem distribuição normal com média m e variância θ^2 . A função densidade de probabilidade é calculada pela Equação 4 [1].

$$f(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left[\frac{\ln(t)-\mu}{\sigma}\right]^2\right) \quad (4)$$

Onde m e θ são, respectivamente a média e o desvio padrão dos logaritmos neperianos. A média (μ) e o desvio padrão (σ) da distribuição lognormal são calculados, respectivamente, pelas Equações 5 e 6 [3].

$$\mu = \exp \left[m + \frac{\theta^2}{2} \right] \quad (5)$$

$$\sigma = \sqrt{\mu^2 (\exp(\theta^2) - 1)} \quad (6)$$

No caso do modelo lognormal, o tempo médio é igual a média da distribuição e o tempo mediano é calculado integrando a Equação 4 [1].

2.3. A distribuição de Weibull

A variável aleatória t (tempo de sobrevivência) tem distribuição de Weibull, se a sua função densidade de probabilidade é calculada pela Equação 7 [4].

$$f(t, \alpha, \beta) = \beta \alpha^\beta t^{\beta-1} \exp(-(\alpha t)^\beta) \quad (7)$$

Onde β é o parâmetro de escala e α é o parâmetro de forma. A média (μ) e o desvio padrão (σ) da distribuição é dada pelas Equações 9 e 10 [4].

$$\mu = \alpha^{-1/\beta} \Gamma \left(1 + \frac{1}{\beta} \right) \quad (9)$$

$$\sigma = \sqrt{\alpha^{-2/\beta} \left(\Gamma \left(1 + \frac{2}{\beta} \right) - \Gamma^2 \left(1 + \frac{1}{\beta} \right) \right)} \quad (10)$$

Onde Γ é o operador da função densidade de probabilidade gama.

O tempo médio é calculado usando a Equação 9. Para calcular o tempo mediano de sobrevivência pela função de Weibull utiliza-se a Equação 11 [1].

$$t_{\text{mediano}} = \frac{\ln(2)^{1/\beta}}{\alpha} \quad (11)$$

2.4. O método bootstrap

O método bootstrap foi introduzido por Efron [5] e, desde então, tem sido profundamente estudado, não apenas em estudos teóricos, como também em várias aplicações. Todavia, a necessidade de manipulação de um número geralmente grande de amostras, a sua operacionalidade somente tornou-se viável com o advento e popularização dos microcomputadores. O método consiste num procedimento estatístico computacionalmente intensivo que permite a avaliação de diversas estatísticas, com base nos dados obtidos da amostra. Sendo assim, ele tem como base a ideia de que o pesquisador pode tratar a sua amostra como se fosse a população que deu origem aos dados e utilizar amostragem com reposição de sua amostra experimental para gerar pseudo-amostras e a partir destas estimar características de interesse de certas estatísticas [2]. Neste caso, a inferência estatística bootstrap tem a finalidade de produzir afirmações sobre uma dada característica da população de interesse, a partir de informações colhidas da amostra.

Vários esquemas diferentes de bootstrap têm sido propostos e muitos deles apresentam bom desempenho em uma ampla variedade de situações. Este método pode ser implementado tanto na estatística não-paramétrica quanto na paramétrica, dependendo apenas do conhecimento do problema. No caso não-paramétrico, reamostra-se os dados com reposição, de acordo com uma distribuição empírica estimada, tendo em vista que, no geral, não se conhece a distribuição subjacente aos dados. No caso paramétrico, quando se tem informação suficiente sobre a forma da distribuição dos dados, a amostra bootstrap é formada realizando-se a amostragem diretamente nessa distribuição com os parâmetros desconhecidos substituídos por estimativas paramétricas.

O processo de reamostragem consiste em gerar amostras a partir da amostra original, cujos dados aleatoriamente retirados (com reposição) são utilizados na formação de cada amostra bootstrap. Dessa forma, todo resultado depende diretamente da amostra original. A distribuição da estatística de interesse aplicada aos valores desse tipo de amostragem, condicional aos dados observados, é definida como a distribuição bootstrap dessa estatística [2]. Operacionalmente, o procedimento bootstrap consiste na reamostragem de mesmo tamanho e com reposição dos dados da amostra original, e cálculo da estatística de interesse para cada reamostra, denominada de pseudovalores [6].

Efron e Tibshirani [2] apresentaram as ideias básicas subjacentes ao método de bootstrap, no âmbito da inferência clássica da estatística, como se segue. Com $x = (x_1, x_2, \dots, x_n)$ amostra aleatória obtida a partir de uma população com função de distribuição desconhecida, F , seja, $\hat{\Theta}(x_1, x_2, \dots, x_n)$, um estimador do parâmetro $\Theta(F)$ que, como se indica, depende naturalmente de F . Seja \hat{F} a função de distribuição empírica associada à amostra obtida, tal que a cada valor observado x_i , onde $(i = 1, 2, \dots, n)$, atribui massa probabilística $\frac{1}{n}$. Então, o valor de \hat{F} é calculado pela Equação 12.

$$\hat{F}_{(n)}(x) = \frac{\left[\sum_{i=1}^n I(x_i \leq x) \right]}{n} \quad (12)$$

Onde:

$\hat{F}_{(n)}(x)$ - o estimador não-paramétrico de máxima verossimilhança de F ;

$I(x_i \leq x)$ - função indicadora.

Uma amostra bootstrap é uma amostra $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ obtida de forma aleatória e com reposição a partir da amostra original $x = (x_1, x_2, \dots, x_n)$, também designada população bootstrap. A notação com asterisco indica que x^* não é um novo conjunto de dados reais x , mas sim uma versão randomizada, ou reamostrada de x . A amostra bootstrap consiste dos correspondentes membros de x , onde: $x_1^* = x_{i1}$, $x_2^* = x_{i2}$, \dots , $x_n^* = x_{in}$. O conjunto $(x_{i1}^*, x_{i2}^*, \dots, x_{in}^*)$ representa a i -ésima amostra de tamanho n com reposição dos dados originais do conjunto $x = (x_1, x_2, \dots, x_n)$.

No método bootstrap, a média amostral calculada é denominada por \bar{x}_i . A cada procedimento de reamostragem do conjunto original $x = (x_1, x_2, \dots, x_n)$, correspondem estimadores, dados por $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. Neste caso, o estimador bootstrap da média da

população é a média aritmética, \bar{x}_B , dos n estimadores \bar{x}_i . Sendo este, um método de reamostragem com reposição, pode-se ter, por exemplo: $x_1^* = x_7$, $x_2^* = x_{10}$, $x_3^* = x_2, \dots$, $x_B^* = x_3$. Portanto, o conjunto de dados reamostrados é constituído de elementos do conjunto dos dados originais $x = (x_1, x_2, \dots, x_n)$, onde alguns não aparecem nenhuma vez, outros aparecem uma vez, outros aparecem duas vezes etc.

Da distribuição $\hat{F}_{(n)}(x)$ obtêm-se B amostras bootstrap de mesmo tamanho n , como apresentada na seqüência abaixo:

$$\begin{aligned} x_1^* &= [x_{11}^*, x_{12}^*, \dots, x_{1n}^*] \\ x_2^* &= [x_{21}^*, x_{22}^*, \dots, x_{2n}^*] \\ &\vdots \\ x_B^* &= [x_{B1}^*, x_{B2}^*, \dots, x_{Bn}^*] \end{aligned}$$

Neste caso, o estimador do desvio padrão da população é calculado pela Equação 13.

$$\hat{s}_B = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x}_i - \bar{x}_B)^2} \quad (13)$$

Especificamente, \bar{x}_i , pode ser substituído pelo estimador $\hat{\Theta}_i$, para cada procedimento de reamostragem. A média \bar{x}_B pode também ser substituída por $\hat{\Theta}_B$, que é a média aritmética dos n estimadores bootstrap. A diferença $\hat{\Theta}_B - \hat{\Theta}_i$ é o estimador do enviesamento de $\hat{\Theta}$. Deste modo, o estimador do erro padrão de $\hat{\Theta}$ é calculado pela Equação 14.

$$\hat{s}_B = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{\Theta}_i - \hat{\Theta}_B)^2} \quad (14)$$

Onde: $\hat{\Theta}_i = \sum_{i=1}^n \frac{x_i^*}{n}$ e $\hat{\Theta}_B = \sum_{i=1}^B \frac{\hat{\Theta}_i}{B}$

A grande vantagem do método “bootstrap” é que ele pode ser aplicado à, praticamente, qualquer estatística $\hat{\Theta}$, não se limitando apenas à média $\hat{\Theta} = \bar{x}$. Isto é muito importante, uma vez que para algumas estatísticas ou não existem fórmulas analíticas ou, quando existem, são difíceis e aproximadas para a estimativa dos seus respectivos erros padrões. A reamostragem bootstrap tenta realizar o que seria desejável realizar na prática: repetir os procedimentos experimentais.

2.5. O algoritmo bootstrap

A técnica de reamostragem bootstrap passa pelo algoritmo de Monte Carlo, onde, um dispositivo gerador de números aleatórios selecionou inteiros i_1, i_2, \dots, i_n , cada um dos quais é igual a algum valor entre 1 e n com probabilidade $\frac{1}{n}$. A amostra formada consiste dos

correspondentes elementos do conjunto original $x = (x_1, x_2, \dots, x_n)$ [5]. Na prática, constrói-se a distribuição bootstrap de F por Monte Carlo com um número de repetições, B , suficientemente grande. Um indicador do tamanho adequado de B , independente do custo computacional, é a qualidade da convergência da estimativa do parâmetro para a estimativa

natural do parâmetro $\hat{\Theta}_B(B \rightarrow \infty) \rightarrow \Theta(F)$, sendo a construção do algoritmo geralmente simples [2]. Sua convergência está garantida pela lei dos Grandes Números, pois, $(x_1^*, x_2^*, \dots, x_n^*)$ nada mais são do que uma amostra de variáveis aleatórias independentes e

identicamente distribuídas com distribuição condicional de $\hat{\Theta}_B$. Assim, quando B tende a infinito, a média amostral $\hat{\Theta}_B$ aproxima-se de Θ [3]. O seguinte algoritmo foi construído pelo método de Monte Carlo para determinar o tempo médio de sobrevivência.

(1) Da amostra experimental, sorteou-se, utilizando um gerador de números aleatórios com probabilidade $\frac{1}{n}$, os n valores com reposição para formar as amostras bootstrap de mesmo tamanho da original.

(2) Computou-se a média aritmética em cada procedimento de reamostragem.

(3) Repetiu-se o passo (2) um número B de vezes obtendo, dessa maneira, B valores da estatística em questão.

(4) Obtiveram-se as B médias para formar a distribuição \hat{F} .

(5) Determinou-se o estimador $\hat{\Theta}_B$ da distribuição \hat{F} .

O valor de $\hat{\Theta}_B$ foi utilizado como o valor do tempo médio de sobrevivência. O procedimento de simulação foi realizado utilizando um programa desenvolvido na linguagem C++, com o gerador de números aleatórios *renu* (Vieira) [7]. A Figura 1 ilustra a operação do programa para construir a distribuição bootstrap pelo algoritmo de Monte Carlo. No presente trabalho, as iterações bootstrap foram realizadas de acordo com o tamanho e a variação nos resultados.

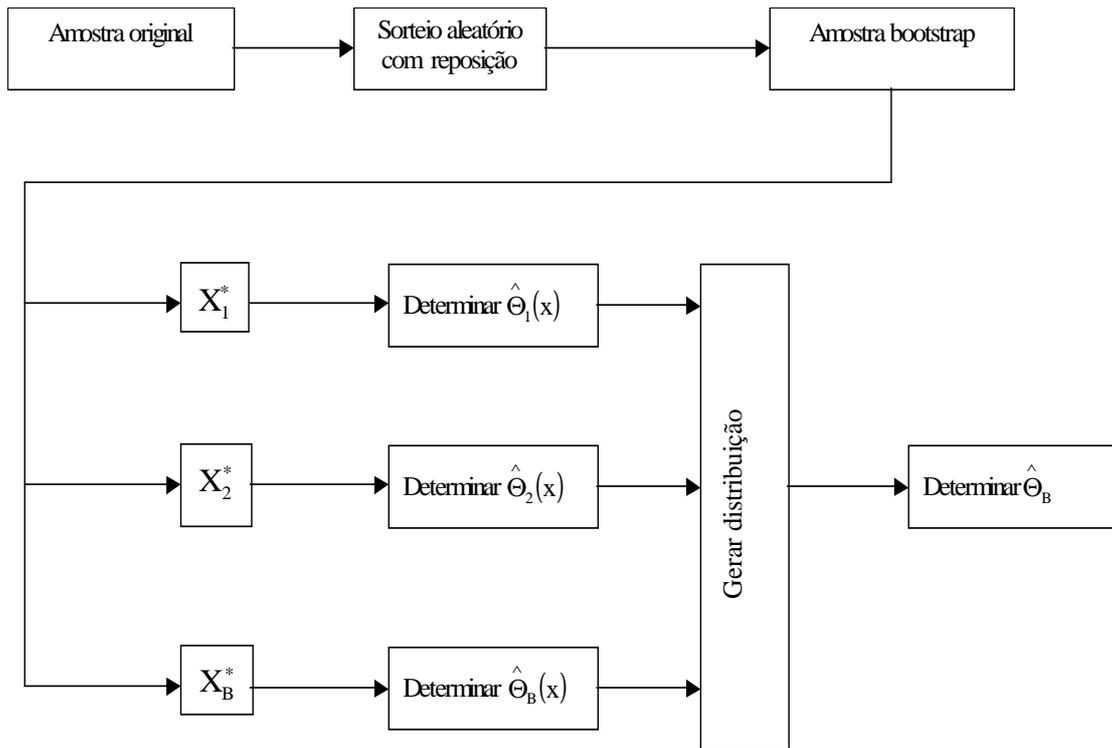


Figura 1: Ilustração esquemática do algoritmo de Monte Carlo para construir a distribuição bootstrap.

3. RESULTADOS E DISCUSSÃO

Para testar o programa computacional desenvolvido utilizaram-se dados de tempo de sobrevivência de 10 pacientes em diálise, obtidos das pesquisas desenvolvidas por Carvalho et al. [1]. Esses dados encontram-se apresentados na Tabela 1.

Tabela 1: Tempo de sobrevivência (em meses) de 10 pacientes em diálise [1].

Paciente	Tempo
1	22
2	6
3	12
4	43
5	23
6	10
7	35
8	18
9	36
10	29

Observando a Tabela 1 pode-se verificar grande variabilidade nos valores do tempo de sobrevida. A Tabela 2 mostra os valores da média aritmética e da mediana dos dados apresentados na Tabela 1.

Tabela 2: Média aritmética e mediana dos tempos de sobrevida (meses).

<i>Estatística</i>	<i>Tempo</i>
Média aritmética \pm desvio padrão	23,4 \pm 12,3
Mediana \pm desvio quartílico	22,5 \pm 23,0

Os resultados apresentados na Tabela 2 mostram elevado grau de dispersão em torno da média aritmética e da mediana. Neste caso, a mediana apresentou o maior valor da medida de dispersão. Assim, a mediana não seria uma medida de tendência central adequada para representar o conjunto de dados da Tabela 1. A Figura 2 mostra o histograma da distribuição dos valores dos tempos de sobrevida apresentados na Tabela 1. Observa-se que a distribuição em questão é bastante achatada. Esta distribuição é do tipo platicúrtica. Numa distribuição deste tipo, os dados obtidos da amostra encontram-se bastante espalhados, dificultando assim, a escolha adequada de uma medida de tendência central que represente corretamente o conjunto de dados. A Figura 3 mostra a representação gráfica da distribuição bootstrap dos valores do tempo de sobrevida para 10.000 iterações.

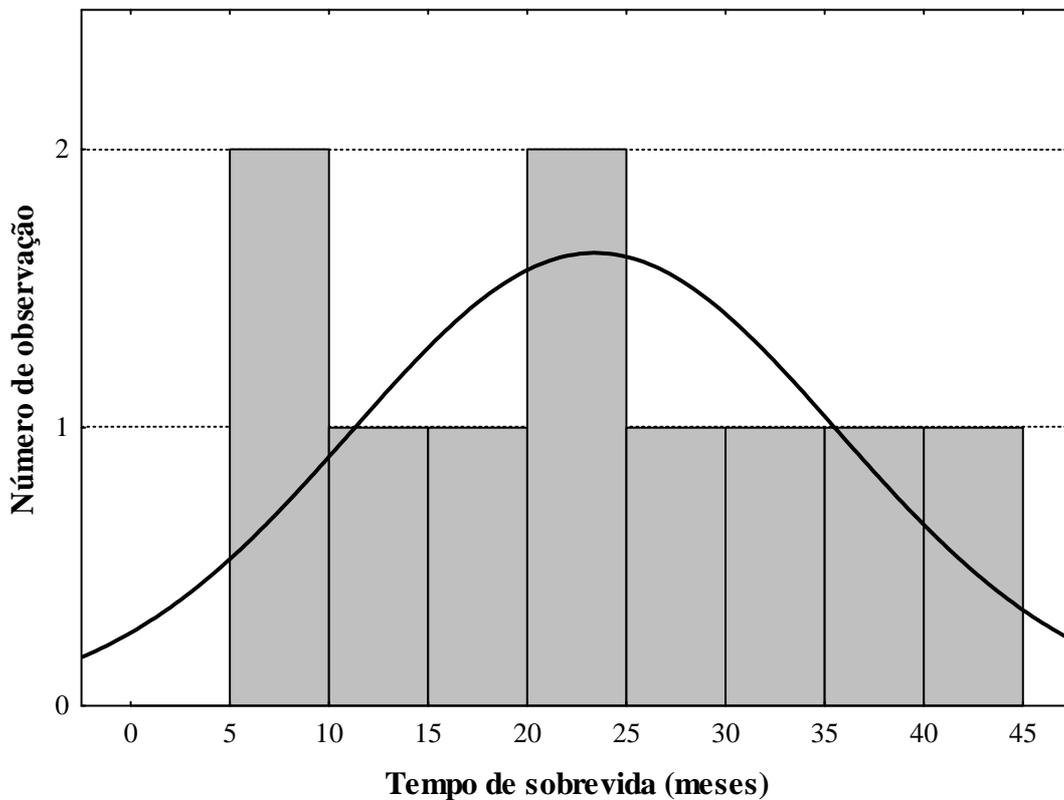


Figura 2: Histograma da distribuição dos tempos de sobrevida.

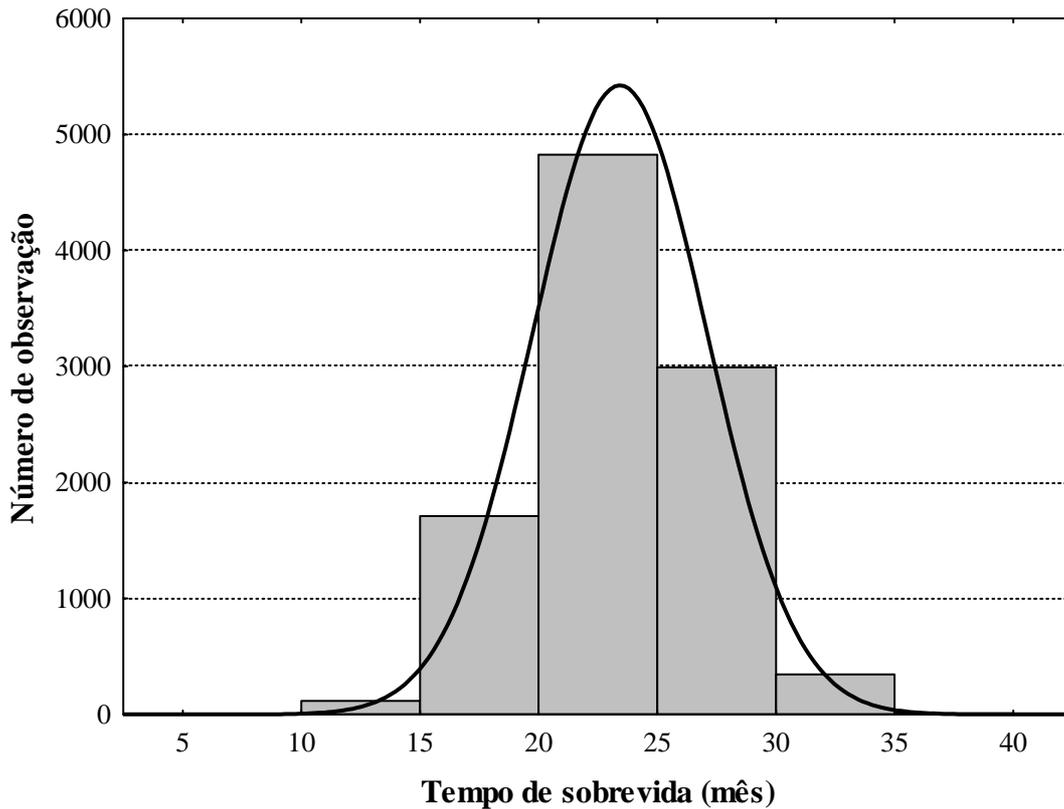


Figura 3: Histograma da distribuição bootstrap dos tempos de sobrevida.

Observa-se que a distribuição bootstrap da Figura 3 é bastante fechada. Este tipo de distribuição é chamado de leptocúrtica. Neste tipo de distribuição os valores encontram-se concentrados próximos do centro da curva. Sendo assim, é possível escolher adequadamente uma medida de tendência central que represente corretamente o conjunto de dados. Na Tabela 3 encontram-se os valores da média aritmética e da mediana dos dados da Tabela 1 analisados pelo método bootstrap.

Tabela 3: Média aritmética e mediana bootstrap dos tempos de sobrevida (meses).

<i>Estatística</i>	<i>Tempo</i>
Média aritmética \pm desvio padrão	23,4 \pm 3,7
Mediana \pm desvio quartílico	23,4 \pm 5,1

Comparando os valores das Tabelas 2 e 3 verifica-se que o método bootstrap reduziu significativamente, a dispersão em torno da média aritmética e da mediana. Isto pode ser observado também, quando se compara os gráficos das Figuras 2 e 3. Por outro lado, não houve diferença significativa entre os valores das médias aritméticas e das medianas nas Tabelas 2 e 3. Isto significa que, a aplicação do método bootstrap na amostra original, não altera os valores da média aritmética e da mediana. Portanto, a média aritmética e a mediana geradas pelo método bootstrap podem ser consideradas medidas de tendência central robustas, pois, segundo, Efron [2], os estimadores bootstrap são robustos para a amostra original.

Segundo Carvalho et al. [1], como a distribuição do tempo de sobrevida é assimétrica, muitas vezes faz mais sentido calcular o tempo mediano de sobrevida ao invés do tempo médio. Por outro lado, como mostra a Tabela 3, é indiferente utilizar o tempo médio e o tempo mediano em

análise de sobrevivência quando as amostras são analisadas pelo método bootstrap. Como na literatura especializada o tempo mediano é mais utilizado em análise de sobrevivência, fez-se necessário realizar comparações entre os valores deste tipo de tempo calculado pela função densidade de probabilidade exponencial, lognormal e Weibull e pelo método bootstrap. Para tanto, utilizou-se os dados da Tabela 1. Os resultados dessas comparações encontram-se apresentadas na Tabela 4.

Tabela 4: Tempo mediano de sobrevivência estimado pela função densidade de probabilidade exponencial, lognormal, Weibull e pelo método bootstrap.

<i>Estatística</i>	<i>Tempo mediano (mês)</i>
Função densidade de probabilidade exponencial	23,4 ± 12,3
Função densidade de probabilidade lognormal	23,4 ± 12,4
Função densidade de probabilidade de Weibull	23,4 ± 12,4
Método bootstrap	23,4 ± 5,1

Conforme apresentado na Tabela 4, não houve diferença significativa entre os valores do tempo mediano. Porém, o valor calculado pelo método bootstrap foi mais preciso, pois, o grau de dispersão foi mais de duas vezes menor do que nas funções de densidade de probabilidade. Assim, o tempo mediano calculado pelo método bootstrap é uma variável robusta que pode ser utilizada como valor mais representativo de conjunto de dados de tempo de sobrevivência.

4. CONCLUSÃO

A média aritmética e a mediana bootstrap foram resistentes aos efeitos causados pelos valores discrepantes, constituindo-se medidas de tendência central que podem ser utilizadas em análise estatística de tempo de sobrevivência.

5. AGRADECIMENTO

Esta pesquisa teve apoio financeiro da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

1. CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; BARBOSA, M. T. S.; SHIMAKURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. Editora Fiocruz, São Paulo, 2005. 396p.
2. EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. Chapman and Hall, New York, 1993. 436 p.
3. AITCHISON, J.; BROWN, J. A. C. *The lognormal distribution*. Cambridge at the University Press, Great Britain. 1957. 534 p.
4. JOHNSON, N. L.; KATZ, S.; BALAKRISHNAN, N. *Continuous univariate distributions*. v. 1. 2. ed. New York: A Wiley-Interscience Publication John Wiley & Sons, INC, 1994. 756 p.
5. EFRON, B. *The jackknife, the bootstrap and other resampling plans*. J.W. Arrowsmith, Ltd., Bristol, 1982. 92 p.
6. MURTEIRA, B. J. F. *Probabilidades e estatística*. McGraw-Hill, Lisboa, 2 ed., 1990. 547p.
7. VIEIRA, J.W. *Construção de um modelo computacional de exposição para cálculos dosimétricos utilizando o código monte carlo egs4 e fantasmas de voxels*. Tese de Doutorado do Programa de Pós-graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, 2004. 146p.