Estimativa de intervalo de confiança para tempo de sobrevida usando desigualdade de Chebyshev via método bootstrap

C. M. Silva[1,2]; R. S. Amaral[2]; J. W. Vieira[2,3]; A. N. C. Silva[2]; J. A. S. Júnior[2]; E. S. Alcoforado[2]

[1]*Departamento de Matemática, Universidade de Pernambuco, 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 09 de abril de 2012)

Quando a amostra é pequena (n<30) e se supõe que a população não é normalmente distribuída, nem a distribuição normal de probabilidade nem a distribuição t de Student podem ser usadas na construção de um intervalo de confiança. Neste caso, utiliza-se a desigualdade de Chebyshev. Na análise estatística de sobrevida, a variável tempo não é normalmente distribuída. Assim, o objetivo do presente trabalho foi utilizar a desigualdade de Chebyshev via método bootstrap para estimar intervalo de confiança para análise de tempo de sobrevida. Para tanto, construiu-se um programa computacional na linguagem C++ com a finalidade de formar a distribuição bootstrap e calcular a média dos tempos de sobrevida. Os resultados mostraram que, a desigualdade de Chebyshev quando usada juntamente com o método bootstrap constituiu-se numa excelente ferramenta estatística para estimar intervalo de tempo de sobrevida.
Palavras-chave: média aritmética; erro padrão; estatística computacional

When the sample is small (n <30) and assumed that the population is not normally distributed, not the normal distribution or the Student t-distribution can be used in the construction of a confidence interval. In this case, we use the Chebyshev inequality. Statistical analysis of survival, the time variable is not normally distributed. The objective of this study was to use the Chebyshev inequality in the bootstrap method to estimate confidence intervals for analysis of survival time. To this end, we constructed a computer program in C ++ in order to form the bootstrap distribution and calculate the average survival times. The results showed that the Chebyshev inequality when used in conjunction with the bootstrap method consisted in an excellent statistical tool for estimating the range of survival time.
Keywords: arithemetic mean; standard error; statistical computing

1. INTRODUÇÃO

Geralmente, os métodos de estimação por intervalo são baseados no pressuposto de que possa ser usada a distribuição normal de probabilidade. Tal pressuposição é garantida sempre que $n \geq 30$, por causa do teorema do limite central, ou n<30, sendo a população normalmente distribuída e o desvio padrão da população (σ) conhecido [1]. Muito embora a média da amostra seja útil como um estimador não tendencioso da média da população, não existe maneira de expressar o grau de acurácia de um estimador por ponto. De fato, matematicamente falando, a probabilidade de que a média da amostra seja exatamente correta como um estimador é P = 0 [1]. Um intervalo de confiança para a média é um intervalo estimado, construído em relação à média da amostra, pelo qual pode ser especificada a probabilidade do intervalo incluir o valor da média da população. O grau de confiança associado com um intervalo de confiança indica a percentagem de tais intervalos incluírem o parâmetro que se está estimando [1].

Os intervalos de confiança para a média são tipicamente construídos com o estimador não tendencioso $\bar{X}$ no centro do intervalo. Os intervalos de confiança mais frequentemente utilizados são os de 90%, 95% e 99% [1].

Se uma população é normalmente distribuída, a distribuição de amostragem da média, para qualquer tamanho da amostra, será também normalmente distribuída, isto é, verdadeiro que o σ

seja conhecido, quer não [1]. Se σ é desconhecido, a expressão $\frac{\left(\bar{X}-\mu\right)}{S_{\bar{X}}}$ inclui no denominador uma variável diferente para cada média da amostra. O resultado é que a inclusão da variável $S_{\bar{X}}$ em lugar da constante $\sigma_{\bar{X}}$ no denominador gera valores convertidos que não se distribuem como valores de Z. Em lugar disso, os valores são distribuídos de acordo com as distribuições t de Student, as quais são platicúrticas (achatadas), se comparadas coma distribuição normal [1].

A distribuição t é apropriada para inferências sobre a média, sempre quando σ for desconhecido e a população normalmente distribuída, qualquer que seja o tamanho da amostra. Contudo, à medida que aumenta o tamanho da amostra, a distribuição t aproxima-se da forma da distribuição normal.

Quando a amostra é pequena (n<30) e se supõe que a população não é normalmente distribuída, nem a distribuição normal de probabilidade, nem a distribuição t de Student podem ser usadas na construção de um intervalo de confiança. Neste caso, um teorema geral desenvolvido pelo matemático russo Chebyshev é bastante útil. Este teorema é conhecido na estatística como desigualdade de Chebyshev. Na verdade, a desigualdade de Chebyshev, é raramente usada na construção de intervalo de confiança para a média, mas é o único método apropriado, quando existe uma população decididamente não normal e uma amostra pequena (n<30) [1].

O tempo de sobrevida não é uma variável que possui distribuição de probabilidade normal. Assim, existem muitas dificuldades para construir adequadamente intervalos de confiança para tempo de sobrevida. Na análise estatística de sobrevida, o mais importante não é saber a ocorrência de um evento de forma pontual, mas o seu desfecho dentro de um intervalo de tempo. Assim, a construção de intervalo de confiança para tempo de sobrevida, constitui um procedimento estatístico importante nas tomadas de decisões.

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 reamostragem bootstrap.

O método 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 do presente estudo foi mostrar o uso da desigualdade de Chebyshev via método bootstrap para estimar intervalo de confiança na análise estatística de sobrevida.

2. MATERIAIS E MÉTODOS

2.1. O teorema de Chebyshev

Segundo Kazmier [1], o teorema de Chebyshev é anunciado da seguinte maneira: a proporção de medidas em um conjunto de dados que se situa dentro de k desvios padrões da média não é menor do que $1-\frac{1}{k^2}$, onde $k \geq 1$.

Aplicando-se a distribuição de amostragem de uma média, a probabilidade de que a média da amostra esteja dentro de k unidades de erro padrão da média da população, é calculado pela Equação 1 [1].

$$P(\bar{X} - \mu / \leq k\sigma_{\bar{X}}) \geq 1 - \frac{1}{k^2} \tag{1}$$

A Equação 1 é geralmente referida como a desigualdade de Chebyshev. Essa equação é baseada na hipótese de que $\sigma_{\bar{X}}$ é conhecido. Se σ é desconhecido, então, pode-se usar $S_{\bar{X}}$ em seu lugar, mas com alguns riscos por causa da flutuação de seu valor em amostras pequenas [1]. Ao usar a desigualdade de Chebyshev na estimação por intervalo, iguala-se $1 - \frac{1}{k^2}$, ao grau de confiança desejado, resolve-se para k, construindo-se, então, o intervalo, utilizando a Equação 2 ou a Equação 3 [1], conforme o caso.

$$\bar{X} \pm k\sigma_{\bar{X}} \tag{2}$$

$$\bar{X} \pm kS_{\bar{X}} \tag{3}$$

O procedimento estatístico utilizado para usar a desigualdade de Chebyshev foi o seguinte: escolheu-se a probabilidade desejada, ou seja, fixou-se antecipadamente a probabilidade de $\bar{X}$ estar no intervalo. Segundo Downing e Clark [3], é comum fixar em 95% essa probabilidade. Assim, calculou-se então, a amplitude do intervalo para que houvesse 95% de chance dele conter o verdadeiro valor de $\bar{X}$. No caso da desigualdade de Chebyshev para o nível de confiança de 95%, o valor de k foi de 4,47 [1].

2.2. O método bootstrap

O método bootstrap foi introduzido por Efron [4] 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 [4]. 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 [4]. 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 [4].

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 [4].

Efron e Tibshirani [2] apresentaram as ideias básicas subjacentes ao método de bootstrap, no âmbito da inferência clássica da estatística, como se segue. Com $x = (x_1, \ x_2, \ \cdots, \ x_n)$ amostra aleatória obtida a partir de uma população com função de distribuição desconhecida, F, seja, $\hat{\Theta}(x_1, \ x_2, \ \cdots \ , x_n)$, 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 $(i = 1, 2, \cdots, n)$, atribui massa probabilística $\frac{1}{n}$. Então, o valor de $\hat{F}$ é calculado pela Equação 4 [2].

$$\hat{F}_{(n)}(x) = \frac{\left[\sum_{i=1}^{n} I(x_i \leq x)\right]}{n} \tag{4}$$

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 $x^* = (x_1^*, \ x_2^*, \ \cdots, \ x_n^*)$ obtida de forma aleatória e com reposição a partir da amostra original $x = (x_1, \ x_2, \ \cdots, \ x_n)$, 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: $x_1^* = x_{i1}, \ x_2^* = x_{i2}, \ \cdots, \ x_n^* = x_{in}$. O conjunto $(x_{i1}^*, x_{i2}^*, \ \ldots \ , \ x_{in}^*)$ representa a i-ésima amostra de tamanho n com reposição dos dados originais do conjunto $x = (x_1, \ x_2, \ \cdots, \ x_n)$ [4].

No método bootstrap, a média amostral calculada é denominada por $\bar{x}_i$. A cada procedimento de reamostragem do conjunto original $x = (x_1, \ x_2, \ \cdots, \ x_n)$, correspondem estimadores, dados por $\bar{x}_1, \ \bar{x}_2, \ \ldots, \ \bar{x}_n$. 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: $x_1^* = x_7, \ x_2^* = x_{10} \ , \ x_3^* = x_2, \cdots, \ x_B^* = x_3$. Portanto, o conjunto de dados reamostrados é constituído de elementos do conjunto dos dados originais $x = (x_1, \ x_2, \ \cdots, \ x_n)$, 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 [4]:

$$x_1^* = \left[x_{11}^*,\ x_{12}^*,\ \cdots,\ x_{1n}^*\right]$$
$$x_2^* = \left[x_{21}^*,\ x_{22}^*,\ \cdots,\ x_{2n}^*\right]$$
$$\vdots$$
$$x_B^* = \left[x_{B1}^*,\ x_{B2}^*,\ \cdots,\ x_{Bn}^*\right]$$

Neste caso, o estimador do desvio padrão da população é calculado pela Equação 5.

$$\hat{S}_B = \sqrt{\frac{1}{n\text{-}1}\sum_{i=1}^{n}(\bar{x}_i - \bar{x}_B)^2} \tag{5}$$

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 6. O erro padrão bootstrap foi calculado pela Equação 7. O intervalo de confiança estimado pela desigualdade de Chebyshev via método bootstrap foi calculada pela Equação 8.

$$\hat{S}_B = \sqrt{\frac{1}{B\text{-}1}\sum_{i=1}^{B}(\hat{\Theta}_i - \hat{\Theta}_B)^2} \tag{6}$$

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

$$S_{\hat{\Theta}_B} = \frac{\hat{S}_B}{\sqrt{n}}. \tag{7}$$

Onde: n é o tamanho da amostra original.

$$\hat{\Theta}_B \pm k S_{\hat{\Theta}_B} = \hat{\Theta}_B \pm k\frac{\hat{S}_B}{\sqrt{n}} \tag{8}$$

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.

2.3. O algoritmo bootstrap

A técnica de reamostragem bootstrap passa pelo algoritmo de Monte Carlo, onde, um dispositivo gerador de números aleatórios selecionou inteiros $i_1,\ i_2,\ \cdots,\ i_n$, 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 $x = (x_1,\ x_2,\ \cdots,\ x_n)$ [4]. 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 $\hat{\Theta}_B(B \to \infty) \to \Theta(F)$, sendo a construção do algoritmo geralmente simples [2]. Sua convergência está garantida pela lei dos Grandes Números, pois, $(x_1^*,\ x_2^*,\ \cdots,\ x_n^*)$ 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 [2]. O seguinte algoritmo foi construído pelo método de Monte Carlo para determinar o tempo médio de sobrevida.

(**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 sobrevida. O procedimento de simulação foi realizado utilizando um programa desenvolvido na linguagem C++, com o gerador de números aleatórios renu (Vieira) [5]. 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.

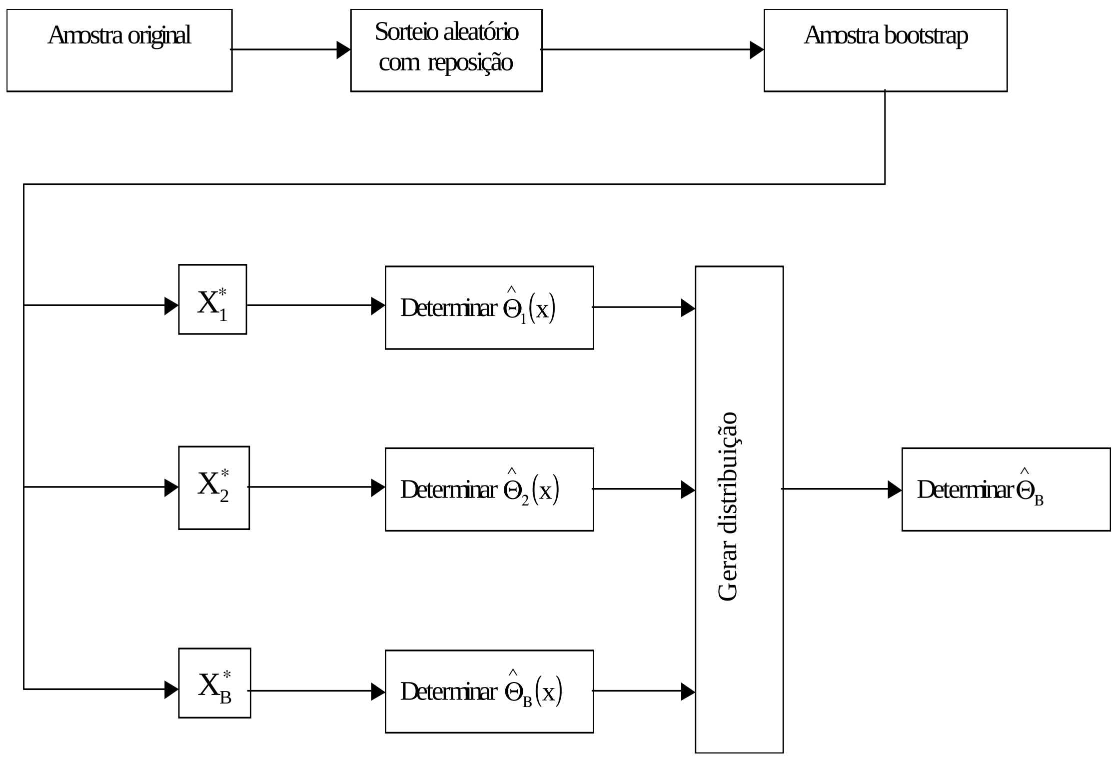

Figura 1: Ilustração esquemática do algoritmo de Monte Carlo para construir a distribuição bootstrap.

3. RESULTADOS E DISCUSSÃO

Para testar o programa computacional desenvolvido utilizaram-se dados de tempo de sobrevida de 21 pacientes com Aids, obtidos das pesquisas desenvolvidas por Carvalho et al. [6]. Esses dados encontram-se apresentados na Tabela 1. Observando a Tabela 1 pode-se verificar grande variabilidade nos valores do tempo de sobrevida. Na Tabela 2 encontram-se os valores da média aritmética e o intervalo de confiança dos dados apresentados na Tabela 1. Os resultados apresentados na Tabela 2 mostram elevado grau de dispersão em torno da média aritmética. 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. Como pode ser observado na Tabela 2, o intervalo de confiança foi bastante amplo, sugerindo elevado grau de dispersão nos valores do tempo de sobrevida. De acordo com Arango [7], o desvio padrão é um fator importante que influência a amplitude do intervalo de confiança. Assim, quanto maior for o desvio padrão, maior será a amplitude do intervalo, como ilustra a Figura 2.

Tabela 1: Tempo de sobrevida (em dias) de 21 pacientes com Aids [1].

Paciente	***Tempo***
1	60
2	82
3	25
4	54
5	80
6	37
7	18
8	29
9	50
10	83
11	80
12	81
13	35
14	52
15	21
16	40
17	22
18	85
19	39
20	16
21	21

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

Estatística	***Tempo***
Média aritmética ± desvio padrão	48,1 ± 25,0
Intervalo da média aritmética	[23,7; 72,5]*

*Com nível de confiança de 95% para a desigualdade de Chebyshev.

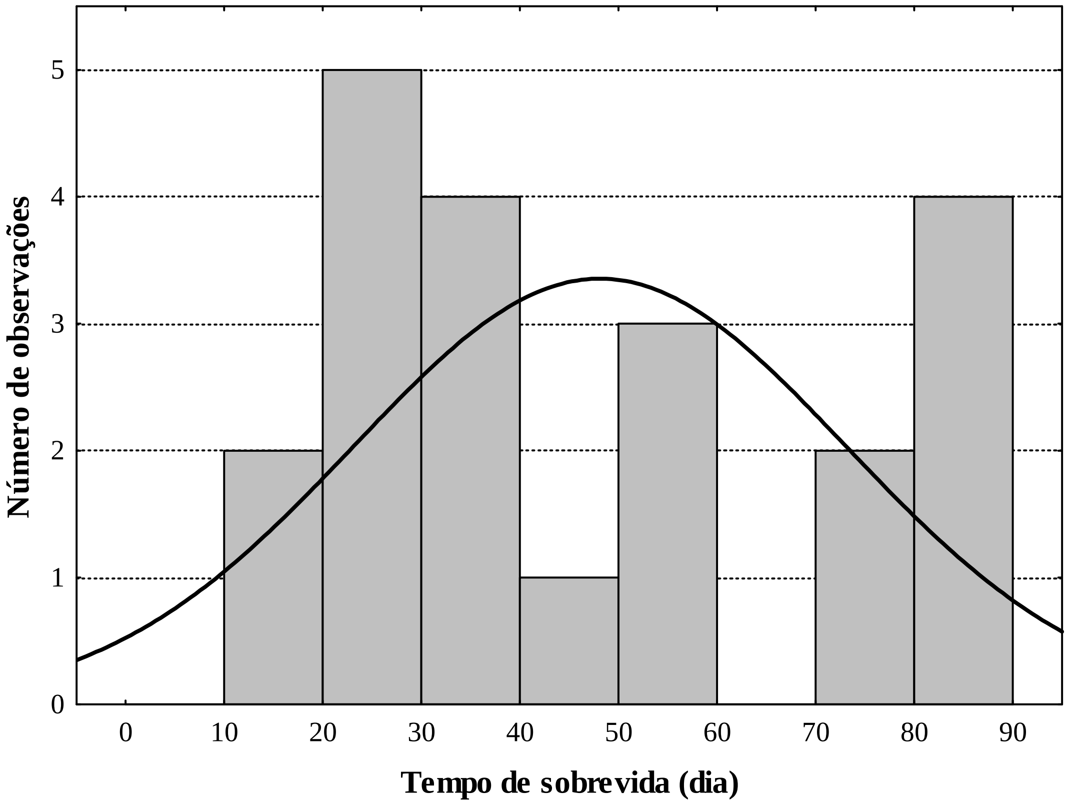

Figura 2: Histograma da distribuição dos tempos de sobrevida.

O método bootstrap quando aplicado num conjunto de dados deduz significativamente a dispersão. A Figura 3 mostra a representação gráfica da distribuição bootstrap dos valores do tempo de sobrevida para 10.000 iterações. 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 apresentados os valores da média aritmética e do intervalo de confiança para o tempo de sobrevida analisados pelo método bootstrap. Comparando os valores das Tabelas 2 e 3 verifica-se que o método bootstrap reduziu significativamente o intervalo de confiança da média aritmética. Segundo Downing e Clark [3], um intervalo de confiança pequeno é melhor porque permite fazer uma estimativa mais precisa do verdadeiro valor de $\overline{X}$. 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 nas Tabelas 2 e 3. Isto significa que, a aplicação do método bootstrap na amostra original, não alterou os valores da média aritmética. Portanto, a média aritmética gerada pelo método bootstrap pode ser considerada medida de tendência central robusta, pois, segundo, Efron [4], os estimadores bootstrap são robustos para a amostra original. Observa-se na Tabela 3 que o desvio padrão foi bastante reduzido como pode ser constatado na Figura 3, pois, ao aplicar o método bootstrap o intervalo de confiança foi deduzido na medida em que aumentou o número de pseudo-amostras.

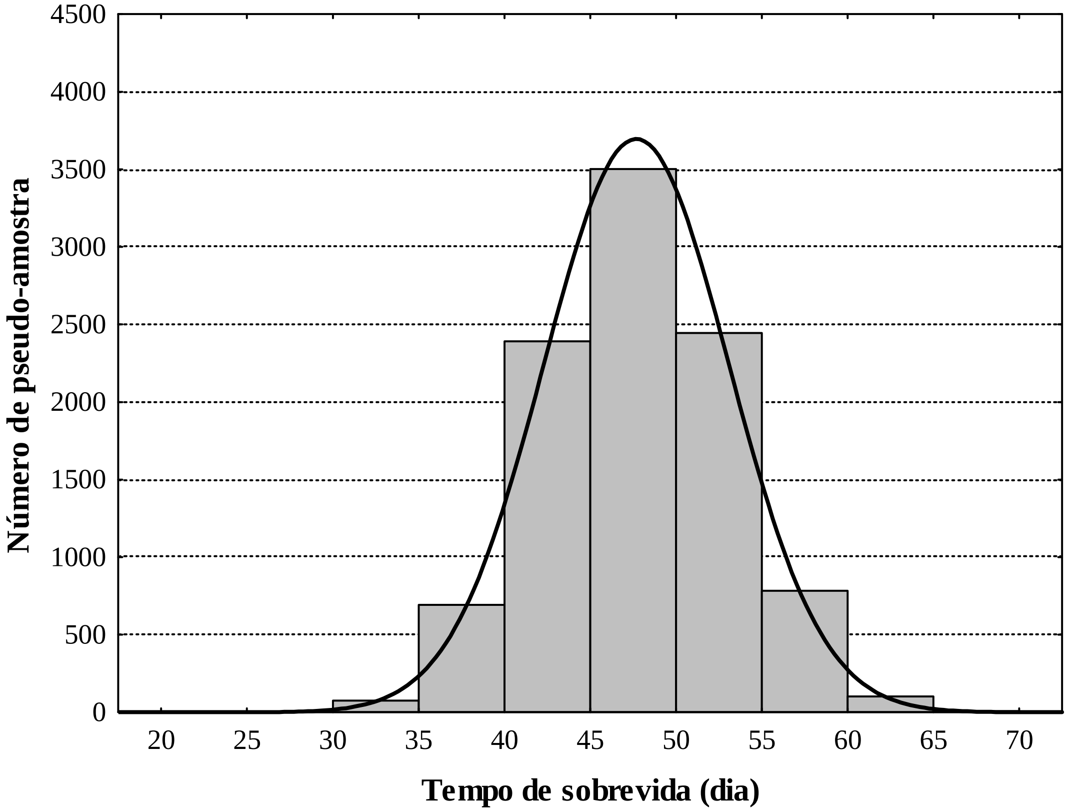

Figura 3: Histograma da distribuição bootstrap dos tempos de sobrevida.

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

Estatística	***Tempo***
Média aritmética ± desvio padrão	47,7 ± 4,3
Intervalo da média aritmética	[43,5; 52,0]*

*Com nível de confiança de 95% para a desigualdade de Chebyshev.

4. CONCLUSÃO

A desigualdade de Chebyshev quando usada com o método bootstrap constituiu um procedimento estatístico preciso na estimativa de intervalo de confiança em análise de tempo de sobrevida.

5. AGRADECIMENTO

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

1. KAZMIER, L. J. Estatística aplicada à economia e administração. McGraw-Hill do Brasil, São Paulo, 1982. 376p.
2. EFRON, B.; TIBSHIRANI, R. J. An introduction to the bootstrap. Chapman and Hall, New York, 1993. 436 p.
3. DOWNING, D.; CLARK, J. Estatística aplicada. Editora Saraiva, São Paulo, 1998. 451p.
4. EFRON, B. The jackknife, the bootstrap and other resampling plans. J.W. Arrowsmith, Ltd., Bristol, 1982. 92 p.
5. VIEIRA, J.W. *Construção de um modelo computacional de exposição para cálculos dosimétricos utilizando o código monte carlo egs4 e fantomas de voxels*. Tese de Doutorado do Programa de Pós-graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, 2004. 146p.
6. CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; BARBOSA, M. T. S.; SHIMAKURA, S. E. Análise de sobrevida: teoria e aplicações em saúde. Editora Fiocruz, São Paulo, 2005. 396p
7. ARANGO, H. G. Bioestatística: teórica e computacional. Guanabara Koogan, Rio de Janeiro, 2 ed., 2005. 547p.