

---

## ESTIMAÇÃO INTERVALAR PARA OS PARÂMETROS DO MODELO EXPONENCIAL DISCRETO: UMA APLICAÇÃO PARA DADOS DE SOBREVIVÊNCIA

CARRASCO, Cleber Giuglioli<sup>1</sup>  
NAKANO, Eduardo Yoshio<sup>2</sup>

---

Recebido em: 2015.08.21

Aprovado em: 2016.03.08

ISSUE DOI: 10.3738/1982.2278.1552

---

**RESUMO:** Este trabalho apresentou a utilização do modelo exponencial discreto (modelo geométrico) na análise de dados discretos, em particular, para dados de sobrevivência. Foram apresentadas a estimativa pontual e seus respectivos intervalos de confiança para o parâmetro do modelo. Este trabalho considerou a técnica de reamostragem *bootstrap* (paramétrica e não-paramétrica) como um procedimento alternativo adequado para construção de intervalos de confiança para o parâmetro do modelo exponencial discreto. A metodologia foi ilustrada através de dados simulados e por meio de uma aplicação em um conjunto de dados reais sobre o tempo até a rehospitalização de pacientes com esquizofrenia. Todos os procedimentos de estimação e simulações foram realizados através do *software* livre R.

**Palavras-chave:** Análise de sobrevivência. *Bootstrap*. Censura. Intervalos de confiança.

**SUMMARY:** This work presents the use of the discrete exponential model (geometric model) in the analysis of discrete data, particularly in survival data. We presented point estimate and their respective confidence intervals for the parameter of the model. This work considered the (parametric and non-parametric) bootstrap resampling techniques as a suitable alternative procedure to construct the confidence intervals for the parameter of discrete exponential model. The methodology was illustrated using simulated data and by an application on a real database about the time until rehospitalization of patients with schizophrenia. All estimation procedures and simulations were performed by the free software R.

**Keywords:** Survival analysis. Bootstrap. Censoring. Confidence intervals.

---

## INTRODUÇÃO

A análise de sobrevivência é um conjunto de técnicas e modelos estatísticos usados na análise de experimentos cuja variável resposta é o tempo até a ocorrência de um evento de interesse. A principal característica desse tipo de análise é a presença de censura, que consiste em uma observação parcial da resposta. Essa informação censurada, apesar de incompleta, é útil e importante para a análise.

Existem situações que os dados de sobrevivência são obtidos de forma discreta devido a imprecisões nas mensurações (ou qualquer outra dificuldade encontrada na coleta desses dados) ou simplesmente por serem discretos. Nesses casos, geralmente são utilizados modelos contínuos para analisar esses dados. Isso ocorre principalmente pela vasta literatura de modelo contínuos e pelo fato dos mesmos estarem disponíveis em muitos *softwares* estatísticos. Entre os autores que apresentam esses modelos podemos destacar Johnson e Kotz (1970), Lawless (1982), Marubini e Valsecchi (1995). Entretanto, nem sempre é aceitável a utilização de um modelo contínuo para a análise de dados discretos, pois em alguns casos pode-se observar um resultado pouco satisfatório. Nakano e Carrasco (2006)

---

<sup>1</sup> Universidade Estadual de Goiás - Câmpus Anápolis de Ciências Exatas e Tecnológicas - Henrique Santillo

<sup>2</sup> Universidade de Brasília - Departamento de Estatística

mostraram que a análise de dados discretos por modelos contínuos apresentam perda na precisão das estimativas, principalmente quando há baixa variabilidade dos valores. Assim, o pesquisador deve estar atento aos seus dados e não utilizar um determinado modelo indiscriminadamente. Outros estudos que consideram modelos discretos para a análise de dados de sobrevivência podem ser vistos em Carrasco; Tutia e Nakano (2012) e Brunello e Nakano (2015).

O objetivo desse trabalho foi apresentar estimativas pontuais e intervalares do parâmetro do modelo exponencial discreto apresentado por Nakano e Carrasco (2006). Os intervalos de confiança do parâmetro do modelo exponencial discreto foram construídos através da teoria assintótica usual, utilizando-se a distribuição normal dos estimadores de máxima verossimilhança (COX; HINKLEY, 1974). Uma vez que, para amostras pequenas ou moderadas, a teoria assintótica usual não pode ser válida (CARRASCO; TUTIA; NAKANO, 2012), as técnicas de reamostragem *bootstrap* paramétrica e não-paramétrica (DAVISON; HINKLEY, 1997) foram utilizadas como procedimentos alternativos para a construção de intervalos de confiança para o parâmetro do modelo exponencial discreto. Foi utilizado o método de Monte Carlo para calcular a probabilidade de cobertura e as amplitudes médias dos intervalos de confiança estudados. A metodologia proposta neste trabalho foi ilustrada em dados artificiais e em um conjunto de dados reais sobre o tempo até a rehospitalização de pacientes com esquizofrenia. Todas as simulações e obtenção das estimativas foram desenvolvidos no *software* livre R (R CORE TEAM, 2015).

## MÉTODOS

### Modelo Exponencial Discreto

Todos os modelos de variáveis contínuas podem ser usados para gerar modelos discretos agrupando os tempos em intervalos unitários. A variável discreta é dada por  $T = [X]$ , onde  $[X]$  representa “a parte inteira de  $X$ ”. A distribuição de probabilidades de  $T$  pode ser escrita como:

$$f(t) = P(T = t) = P(t \leq X < t+1), \quad t = 0, 1, 2, \dots, \quad (1)$$

Como visto em Nakano e Carrasco (2006), no caso em que  $X$  segue a distribuição exponencial com parâmetro  $\lambda$ , tem-se que a função (densidade) de probabilidade, de sobrevivência e de risco da variável aleatória  $T$  são dadas, respectivamente, por:

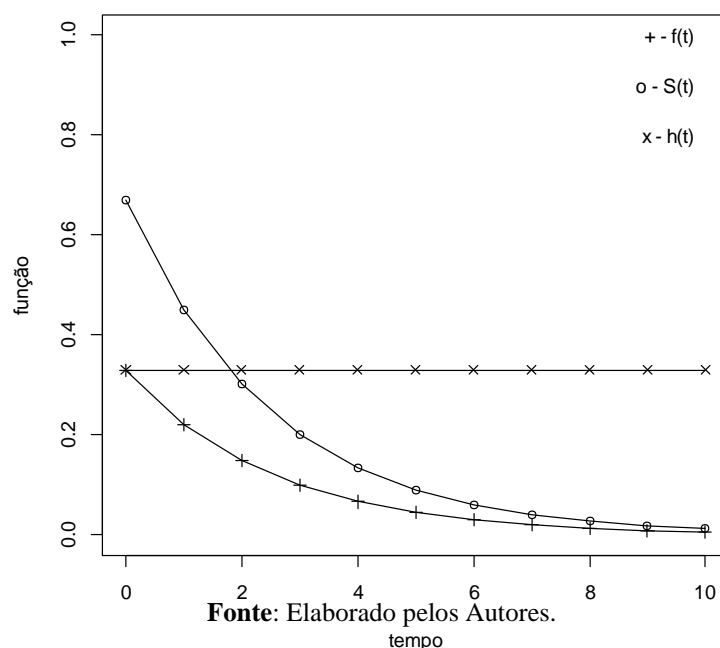
$$f(t | \lambda) = e^{-\lambda t} (1 - e^{-\lambda}), \quad t = 0, 1, 2, \dots \quad (2)$$

$$S(t | \lambda) = P(T > t) = e^{-\lambda(t+1)}, \quad t = 0, 1, 2, \dots \quad (3)$$

$$h(t | \lambda) = 1 - e^{-\lambda}, \quad t = 0, 1, 2, \dots \quad (4)$$

A Figura 1 apresenta as funções de probabilidade (2), sobrevivência (3) e de risco (4) do modelo exponencial discreto para  $\lambda = 0,40$ .

**Figura 1:** Função de probabilidade, sobrevivência e de risco do modelo exponencial discreto.



### Estimador de Máxima Verossimilhança

Assumindo que a contribuição para a verossimilhança do tempo censurado em  $t$ , seja  $S(t) = P(T > t)$  (KALBFLEISCH e PRENTICE, 1980), tem-se que, para o modelo exponencial discreto (2) e (3), a função de verossimilhança apresenta a seguinte forma:

$$L(\lambda | \mathbf{t}, \boldsymbol{\delta}) = \prod_{i=1}^n (1 - e^{-\lambda})^{\delta_i} e^{-\lambda(t_i + 1 - \delta_i)} \quad (5)$$

em que,  $\lambda$  é o parâmetro a ser estimado,  $t = (t_1, \dots, t_n)$  é o vetor dos valores observados, com seus respectivos indicadores de censuras dados por  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ .

Neste caso, o Estimador de Máxima Verossimilhança (EMV) do parâmetro  $\lambda$  do modelo exponencial discreto é dado por (NAKANO; CARRASCO, 2006):

$$\hat{\lambda} = \ln \left( \frac{\sum_{i=1}^n t_i + n}{\sum_{i=1}^n t_i + n - \sum_{i=1}^n \delta_i} \right). \quad (6)$$

### Estimação Intervalar

Um intervalo de confiança para o parâmetro do modelo exponencial discreto pode ser baseado na teoria assintótica usual, utilizando-se a distribuição normal dos estimadores de máxima verossimilhança (COX e HINKLEY, 1974). Assim, um intervalo de  $100(1 - \alpha)\%$  de confiança para o parâmetro  $\lambda$  é dado por:

$$IC(\lambda, (1 - \alpha)\%) = \hat{\lambda} \pm z_{(\alpha/2)} \sqrt{\frac{\sum_{i=1}^n \delta_i}{\left( \sum_{i=1}^n t_i + n \right) \left( \sum_{i=1}^n t_i + n - \sum_{i=1}^n \delta_i \right)}}. \quad (7)$$

Onde  $\hat{\lambda}$  é o EMV do parâmetro  $\lambda$ , dado por (6), e  $z_{(\alpha/2)}$  é o quantil  $\alpha/2$  de uma distribuição normal padrão.

A utilização do intervalo de confiança (7) deve ser direcionada pelo tamanho da amostra, a qual deve ser suficientemente grande (LOUZADA-NETO; PERDONÁ; CARRASCO, 2009). No entanto, na análise de sobrevivência é comum a presença de amostras pequenas ou moderadas no qual a utilização de (7) pode não ser válida. Nesses casos, as técnicas de reamostragem *bootstrap* paramétrica e não-paramétrica (DAVISON e HINKLEY, 1997) podem ser utilizadas como uma alternativa na construção de intervalos de confiança para o parâmetro do modelo exponencial discreto. Essa técnica foi primeiramente proposta por Efron (1979) e visa reamostrar a amostra original.

Dessa forma, para cada reamostragem da amostra original calcula-se o EMV para  $\square$  e tem-se no final de  $B$  reamostragens,  $\hat{\lambda}_1 < \dots < \hat{\lambda}_B$  valores dos EMV ordenados. Utiliza-se então:

$$\hat{\lambda}_{(B)\left(\frac{\alpha}{2}\right)} \text{ e } \hat{\lambda}_{(B)\left(1-\frac{\alpha}{2}\right)} \quad (8)$$

como os limites inferiores e superiores do intervalo  $100(1 - \alpha)\%$  de confiança para  $\square$ , em que  $\alpha$  é o nível de significância. Neste trabalho utilizou-se  $B = 1.000$ .

Se as reamostras são obtidas através da reposição dos elementos da amostra original tem-se o *bootstrap* não-paramétrico. Se as reamostras forem geradas através do modelo ajustado com os valores dos parâmetros fixados nos EMV obtidos na amostra original tem-se o *bootstrap* paramétrico (LOUZADA-NETO; PERDONÁ; CARRASCO, 2009).

### Estudo de Monte Carlo

A probabilidade de cobertura de um intervalo de confiança é determinada repetindo a construção do intervalo de confiança  $D$  vezes, nas quais se verifica em cada repetição se o valor do parâmetro pertence ou não a esse intervalo. Dessa forma, a probabilidade de cobertura para um intervalo de confiança pode ser calculada por (CARRASCO; TUTIA; NAKANO, 2012):

$$1 - \frac{\sum_{i=1}^D \psi(vp \notin IC_i)}{D} \quad (9)$$

Onde  $\psi(\cdot)$  é uma função indicadora que vale 1 se  $vp \notin IC_i$  e 0 caso contrário,  $vp$  é o valor do parâmetro e  $IC_i$  é o  $i$ -ésimo intervalo de confiança construído,  $i = 1, 2, \dots, D$ . Neste trabalho utilizou-se  $D = 1.000$ .

Tanto a probabilidade de cobertura quanto a amplitude média de um intervalo de confiança são critérios para comparação de procedimentos para construção de intervalos de confiança. A amplitude média de um intervalo de confiança pode ser calculada de maneira similar a probabilidade de cobertura: repete-se o procedimento de construção do intervalo de confiança  $D$  vezes; em cada repetição calcula-se a amplitude desse intervalo e no final das  $D$  repetições calcula-se a média dessas amplitudes. A forma de cálculo da probabilidade de cobertura e da amplitude média justifica o uso do método de Monte Carlo (HAMMERSLEY; HANDSCOMB, 1964), uma vez que, esse método consiste em gerar, a partir da distribuição dos dados, um número muito grande de amostras e, em cada uma delas, obter o intervalo de confiança. Assim, a probabilidade de cobertura é a razão do número de vezes que o intervalo de confiança contém o verdadeiro valor do parâmetro utilizado na geração das amostras sobre o número total de simulações realizadas.

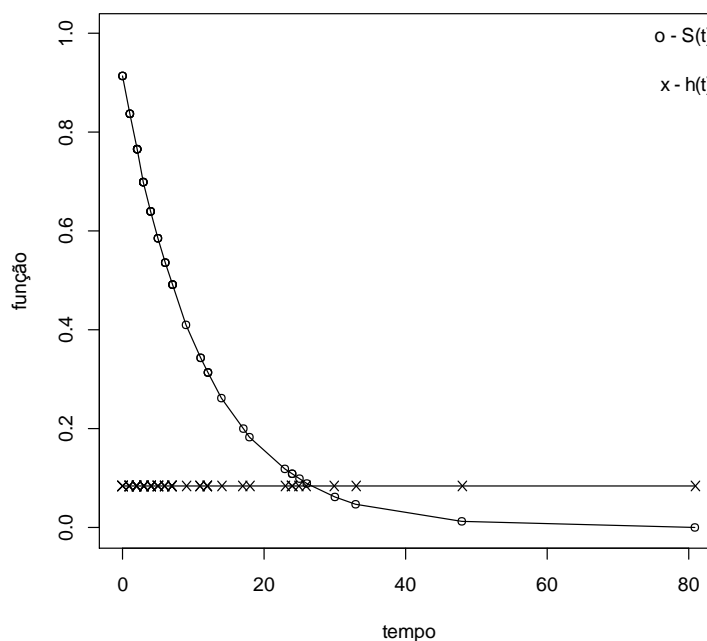
## RESULTADO E DISCUSSÃO

Para gerar dados do modelo exponencial discreto utilizou-se a distribuição geométrica, uma vez que, pode-se reescrever o modelo (2) como sendo uma geométrica de parâmetro  $\theta = 1 - e^{-\lambda}$ , tal que  $f(t | \theta) = \theta(1 - \theta)^t$ ,  $t = 0, 1, 2, \dots$ . Dessa forma, geraram-se conjuntos de dados com  $\square$  fixado em 0,10, de tamanho  $n = 20, 50$  e  $200$ , com 0%, 10% e 30% de dados censurados. O mecanismo de censura considerado foi o aleatório e as censuras foram geradas através de uma distribuição Bernoulli (CASELLA; BERGER, 2010). A escolha desse modelo permite controlar o percentual de dados censurados na amostra através da probabilidade de sucesso da distribuição Bernoulli. Ao todo foram gerados nove conjuntos de dados, combinando os diversos tamanhos de amostra e percentuais de censura. Todo o procedimento foi implementado no *software* R (R CORE TEAM, 2015). A escolha da utilização do *software* R neste trabalho é que, além de ser gratuito, apresenta código-fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos pelo usuário (MELLO; PETERNELLI, 2013).

Para a amostra gerada com  $\square = 0,10$ ,  $n = 50$  e sem censura, os dados simulados foram: 3, 7, 3, 3, 33, 18, 4, 5, 9, 23, 25, 26, 11, 4, 2, 0, 0, 1, 12, 48, 1, 30, 5, 17, 7, 11, 3, 3, 4, 1, 2, 12, 2, 6, 12, 1, 81, 7, 24, 6, 2, 6, 7, 5, 4, 14, 2, 0, 24, 2. Para esse conjunto de dados estimou-se o parâmetro do modelo exponencial discreto em  $\hat{\lambda} = 0,0889$ .

A Figura 2 apresenta as funções de sobrevivência (3) e de risco (4) do modelo exponencial discreto estimadas a partir dos dados simulados.

**Figura 2:** Função estimada de sobrevivência e de risco do modelo exponencial discreto para a amostra de  $n = 50$ ,  $\square = 0,10$  e 0% de censura.



**Fonte:** Elaborado pelos Autores.

A Tabela 1 apresenta os resultados dos intervalos de confiança assintótico, *bootstrap* percentil paramétrico e não paramétrico, juntamente com os vícios dos estimadores obtidos via *bootstrap* paramétrico e não paramétrico, que foram calculados da seguinte forma:

$$\text{média} \left( \frac{(\hat{\lambda}_1^* - \hat{\lambda})}{\hat{\lambda}} \right) \text{ e } \text{média} \left( \frac{(\hat{\lambda}_2^* - \hat{\lambda})}{\hat{\lambda}} \right) \quad (10)$$

em que  $\hat{\lambda}_1^*$  e  $\hat{\lambda}_2^*$  são os vetores de estimativas *bootstrap* paramétrico e não paramétrico, respectivamente de tamanho  $B = 1.000$ , e  $\hat{\lambda}$  é o EMV de  $\lambda$ , obtido por (6).

Observa-se na Tabela 1 que os intervalos de confiança obtidos pelos diferentes procedimentos estão próximos e que, para esta amostra, o intervalo de confiança via técnica *bootstrap* não paramétrico obteve amplitude e vício maior do que os demais.

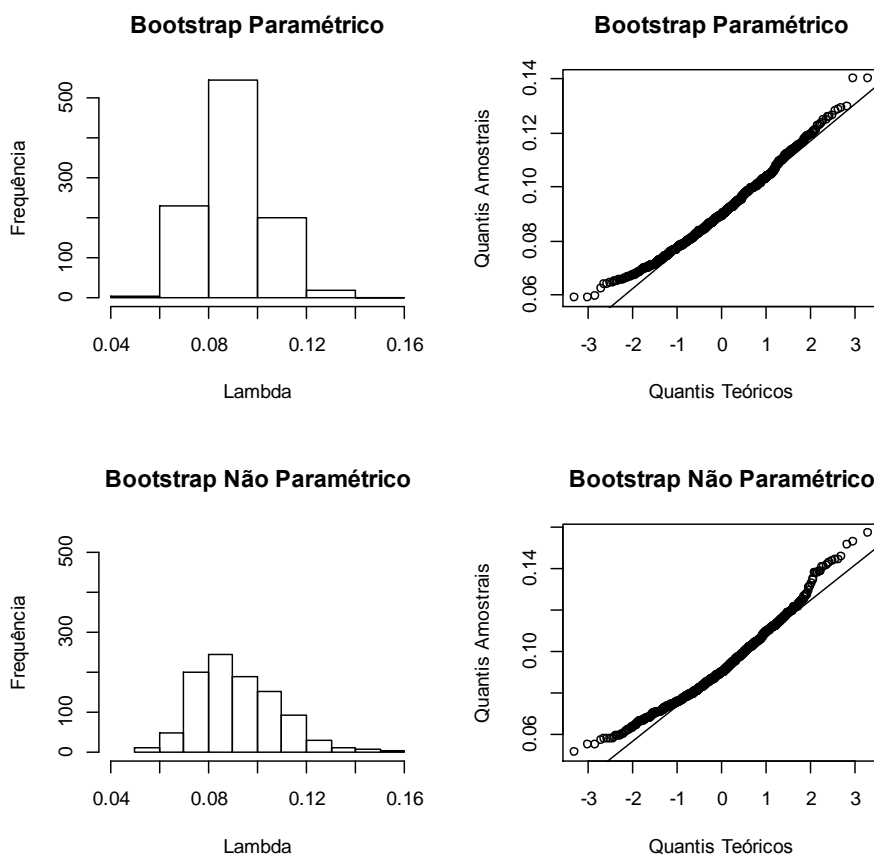
**Tabela 1:** Intervalos de Confiança para o parâmetro do modelo exponencial discreto para a amostra de  $n = 50$ ,  $\lambda = 0,10$  e 0% de censura.

Intervalo de Confiança	IC ( $\lambda$ , 95%)	Amplitude	Vício
Assintótico	[0,0642 ; 0,1135]	0,0493	-----
Bootstrap Paramétrico	[0,0671 ; 0,1186]	0,0515	0,0157
Bootstrap Não Paramétrico	[0,0640 ; 0,1294]	0,0654	0,0342

**Fonte:** Elaborado pelos Autores.

A Figura 3 apresenta os histogramas e os *qq-plots* das distribuições empíricas do estimador de máxima verossimilhança obtidos via técnica *bootstrap* paramétrica e não-paramétrica para os dados simulados. Note que há um indicativo de não normalidade do estimador, sugerindo, neste caso, que a teoria usual de verossimilhança pode não propiciar resultados suficientemente adequados. O teste de Shapiro Wilk (SHAPIRO e WILK, 1965) confirma o indicativo dos histogramas e *qq-plots* da Figura 3. As estatísticas  $W$  de Shapiro Wilk são iguais a 0,9881 e 0,9765 para as distribuições obtidas via *bootstrap* paramétrico e não paramétrico, respectivamente. Assim, a um nível de significância de 1% rejeita-se a normalidade do estimador de máxima verossimilhança para ambos os casos.

**Figura 3:** *QQ-plots* e histogramas das distribuições empíricas do estimador do parâmetro do modelo exponencial discreto via *bootstrap* paramétrico e não paramétrico.



**Fonte:** Elaborado pelos Autores.

As Tabelas 2, 3 e 4 apresentam os resultados do estudo de simulação de Monte Carlo com as estimativas de máxima verossimilhança para  $\lambda = 0,10$ , probabilidades de cobertura, amplitudes médias e os vícios médios dos intervalos de confiança assintótico, *bootstrap* percentil paramétrico e não paramétrico, para as amostras com 0%, 10% e 30% de dados censurados, respectivamente.

Observa-se nas Tabelas 2, 3 e 4 que as estimativas de máxima verossimilhança se aproximam do verdadeiro valor do parâmetro a medida que o tamanho da amostra aumenta e, quanto maior o número de dados censurados menor é a estimativa. Sem dados censurados as probabilidades de cobertura dos intervalos ficaram próxima da nominal de 95%, exceto para os intervalos de confiança *bootstrap* não paramétrico, que para as amostras de tamanho 20 e 50 ficaram abaixo desse valor. Com dados censurados as probabilidades de cobertura de todos os intervalos de confiança ficaram bem abaixo de 95%, principalmente quando se aumenta o tamanho da amostra. Nesses casos o intervalo *bootstrap* paramétrico foi o que apresentou menor probabilidade de cobertura. Isso se deve ao fato de que quanto maior o percentual de censura, mais a amostra censurada se distancia da distribuição gerada. Com relação as amplitudes médias, quanto maior o tamanho amostral menor as amplitudes dos intervalos, isso também ocorre para os dados censurados. Os intervalos de confiança *bootstrap* paramétricos obtiveram maiores vícios médios em relação aos intervalos não paramétricos. Na presença de dados censurados os vícios médios dos intervalos *bootstrap* não paramétricos diminuíram, enquanto para os intervalos *bootstrap* paramétricos os vícios aumentaram e foram negativos.

**Tabela 2:** Probabilidade de cobertura, amplitude média e vício para amostras sem censuras.

$n$ e EMV	IC ( $\lambda$ , 95%)	Prob. de Cobertura	Amplitude Média	Vício Médio
$n = 20$ $\hat{\lambda} = 0,0810$	Assintótico	0,950	0,0921	-----
	<i>Bootstrap</i> Paramétrico	0,942	0,1014	0,0528
	<i>Bootstrap</i> N-P	0,904	0,0940	0,0476
$n = 50$ $\hat{\lambda} = 0,0889$	Assintótico	0,942	0,0563	-----
	<i>Bootstrap</i> Paramétrico	0,941	0,0584	0,0204
	<i>Bootstrap</i> N-P	0,925	0,0561	0,0194
$n = 200$ $\hat{\lambda} = 0,1101$	Assintótico	0,955	0,0278	-----
	<i>Bootstrap</i> Paramétrico	0,953	0,0281	0,0053
	<i>Bootstrap</i> N-P	0,950	0,0279	0,0050

Fonte: Elaborado pelos Autores.

**Tabela 3:** Probabilidade de cobertura, amplitude média e vício para amostras com 10% de dados censuras.

$n$ e EMV	IC ( $\lambda$ , 95%)	Prob. de Cobertura	Amplitude Média	Vício Médio
$n = 20$ $\hat{\lambda} = 0,0601$	Assintótico	0,886	0,0869	-----
	<i>Bootstrap</i> Paramétrico	0,897	0,0856	-0,0574
	<i>Bootstrap</i> N-P	0,887	0,0837	0,0470
$n = 50$ $\hat{\lambda} = 0,0759$	Assintótico	0,843	0,0530	-----
	<i>Bootstrap</i> Paramétrico	0,727	0,0492	-0,0858
	<i>Bootstrap</i> N-P	0,872	0,0498	0,0192
$n = 200$ $\hat{\lambda} = 0,1014$	Assintótico	0,640	0,0263	-----
	<i>Bootstrap</i> Paramétrico	0,180	0,0237	-0,0993
	<i>Bootstrap</i> N-P	0,662	0,0248	0,0050

Fonte: Elaborado pelos Autores.

**Tabela 4:** Probabilidade de cobertura, amplitude média e vício para amostras com 30% de dados censuras.

$n$ e EMV	IC ( $\lambda$ , 95%)	Prob. de Cobertura	Amplitude Média	Vício Médio
$n = 20$ $\hat{\lambda} = 0,0437$	Assintótico	0,611	0,0757	-----
	<i>Bootstrap</i> Paramétrico	0,263	0,0573	-0,2722
	<i>Bootstrap</i> N-P	0,625	0,0635	0,0458
$n = 50$ $\hat{\lambda} = 0,0650$	Assintótico	0,317	0,0463	-----
	<i>Bootstrap</i> Paramétrico	0,007	0,0332	-0,2934
	<i>Bootstrap</i> N-P	0,307	0,0380	0,0187
$n = 200$ $\hat{\lambda} = 0,0763$	Assintótico	0,004	0,0229	-----
	<i>Bootstrap</i> Paramétrico	0,000	0,0160	-0,3063
	<i>Bootstrap</i> N-P	0,003	0,0189	0,0049

Fonte: Elaborado pelos Autores.

### Aplicação em Dados Reais

A seguir tem-se uma aplicação do uso do modelo exponencial discreto a um conjunto de dados reais retirados de Werneck *et al.* (2011), que foram posteriormente analisados, dentro de um contexto bayesiano, por Fernandes (2014). Os dados são sobre o tempo até a rehospitalização (readmissão do paciente por motivos psiquiátricos) de pacientes com esquizofrenia e que fazem o uso do medicamento antipsicótico Risperidona. Os dados foram baseados em uma amostra de 63 pacientes do Instituto de Psiquiatria da Universidade de São Paulo. O tempo até a rehospitalização foi medido em meses, onde  $t = 0$  indica que o paciente retornou ao hospital menos de um mês após sua última visita. Os dados estão condensados na Tabela 5.



**Tabela 5:** Tempo até a rehospitalização de pacientes diagnosticados com esquizofrenia e que fazem uso do medicamento Risperidona.

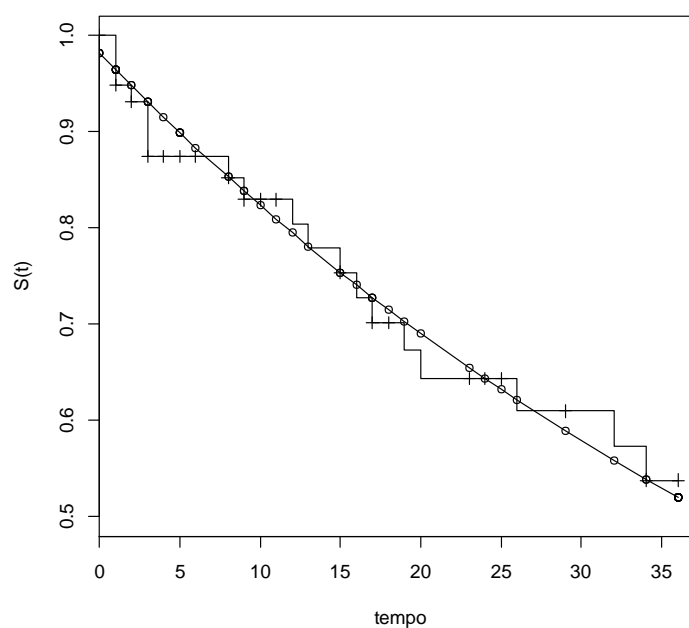
Tempo até rehospitalização (meses*)	Nº de indivíduos sob risco no início do mês	Nº de rehospitalizações no mês	Nº de censuras no mês
0	63	0	4
1	59	3	4
2	52	1	1
3	50	3	8
8	39	1	1
9	37	1	3
12	33	1	0
13	32	1	0
15	31	1	1
16	29	1	0
17	28	1	3
19	24	1	0
20	23	1	3
26	19	1	1
32	17	1	0
34	16	1	15

\*Os tempos originais são em dias e aqui foram transformados em meses para ilustrar o modelo discreto.

**Fonte:** Werneck *et al.* (2011) e Fernandes (2014).

Para os dados da Tabela 5 o parâmetro do modelo exponencial discreto foi estimado em  $\hat{\lambda} = 0,0177$ . A Figura 4 apresenta as curvas de sobrevivência do modelo exponencial discreto juntamente com a curva de Kaplan-Meier. Note que o modelo exponencial discreto se ajusta bem aos dados.

**Figura 4:** Curva de Sobrevivência de Kaplan-Meier (escada) e do modelo exponencial discreto estimadas para os dados de Werneck *et al.* (2011).



**Fonte:** Elaborado pelos Autores.

Os resultados dos intervalos de confiança assintótico, *bootstrap* paramétrico e não paramétrico para os dados de Werneck *et al.* (2011) juntamente com as suas amplitudes e vícios estão condensados na Tabela 6. As amplitudes dos intervalos via técnica *bootstrap* foram menores do que o assintótico para esse exemplo e, o vício do intervalo *bootstrap* paramétrico foi maior do que o não paramétrico. Observe que os resultados encontrados na Tabela 6 estão de acordo com o estudo de simulação de Monte Carlo apresentado nas Tabelas 2, 3 e 4.

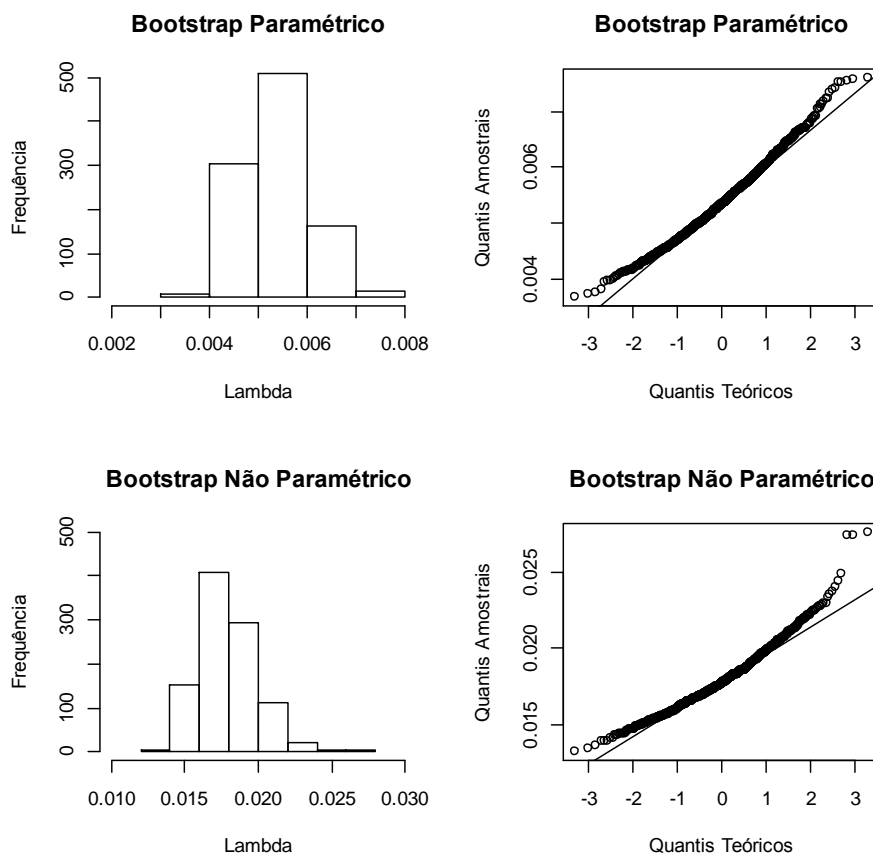
**Tabela 6:** Intervalos de Confiança para o parâmetro do modelo exponencial discreto para o dados da Tabela 5.

Intervalo de Confiança	IC ( $\lambda$ , 95%)	Amplitude	Vício
Assintótico	[0,0097 ; 0,0256]	0,0159	-----
Bootstrap Paramétrico	[0,0042 ; 0,0068]	0,0026	-0,6957
Bootstrap Não Paramétrico	[0,0147 ; 0,0222]	0,0075	0,0140

Fonte: Elaborado pelos Autores.

Os histogramas e os *qq-plots* das distribuições empíricas do estimador de máxima verossimilhança obtidos via técnica *bootstrap* paramétrica e não paramétrica para os dados de Werneck *et al.* (2011) estão apresentados na Figura 5. Note que há um indicativo de não normalidade do estimador, o que é confirmado pelo teste de Shapiro Wilk, que apresenta a estatística  $W$  igual a 0,9890 e 0,9687 para as distribuições obtidas via *bootstrap* paramétrico e não paramétrico, respectivamente. Ou seja, a um nível de significância de 1% rejeita-se a normalidade do estimador de máxima verossimilhança para ambos os casos.

**Figura 5:** *QQ-plots* e histogramas das distribuições empíricas do estimador do parâmetro do modelo exponencial discreto via *bootstrap* paramétrico e não paramétrico para os dados de Werneck *et al.* (2011).



Fonte: Elaborado pelos Autores.

## CONCLUSÃO

O modelo exponencial discreto pode ser utilizado para modelagem de dados discretos de sobrevivência evitando o uso de modelos contínuos para tratar esse tipo de dados, uma vez que, o uso inadequado de modelos contínuos para análise de dados discretos pode acarretar em estimativas não satisfatórias, havendo possível perda de precisão das mesmas. Dessa forma, é necessário cuidado e atenção na escolha do modelo a ser utilizado para análise dos dados. A técnica de reamostragem *bootstrap* proporcionou verificar que a normalidade dos estimadores de máxima verossimilhança neste caso não é válida, assim, os intervalos de confiança *bootstrap* tanto paramétrico quanto não paramétrico podem ser utilizados como procedimento alternativo na estimação intervalar do parâmetro do modelo exponencial discreto. Em relação ao estudo de simulação de Monte Carlo, as probabilidades de cobertura dos intervalos de confiança para as amostras sem censura foram satisfatórias. No entanto, essas probabilidades de cobertura foram baixas na presença de censuras. Esse fato possivelmente ocorreu porque a distribuição dos dados censurados não segue exatamente a distribuição considerada para gerar os dados não censurados. Quanto às amplitudes médias, os intervalos de confiança assintóticos e via técnica *bootstrap* tanto paramétrico quanto não paramétrico tiveram o mesmo comportamento: as amplitudes diminuíram com o aumento do tamanho da amostra e/ou percentual de censuras.

A facilidade computacional de implementação dos métodos de estimação pontual e intervalar do parâmetro do modelo exponencial discreto e a utilização do *software* R que, por ser gratuito e apresentar código-fonte aberto, pode encorajar a utilização dos modelos discretos na análise de sobrevivência.

## REFERÊNCIAS

- BRUNELLO, G. H. V.; NAKANO, E. Y. Inferência bayesiana no modelo Weibull discreto em dados com presença de censura. **TEMA - Tend. Mat. Apl. Comput.**, v.16, n.2, p.97-110, 2015.
- CARRASCO, C.G.; TUTIA, M.H.; NAKANO, E.Y. Intervalos de Confiança para os Parâmetros do Modelo Geométrico com Inflação de Zeros. **TEMA - Tend. Mat. Apl. Comput.**, v. 13, n.3, p.247-255, 2012.
- CASELLA, G.; BERGER, R.L. **Inferência Estatística**. Tradução da 2ª Edição Norte-Americana. Editora: Cengage Learning, 2010.
- COX, D.R.; HINKLEY, D.V. **Theoretical statistics**. London: Chapman and Hall, 1974.
- DAVISON, A.C.; HINKLEY, D.V. **Bootstrap methods and their application**. Cambridge: Cambridge University Press, 1997.
- EFRON, B. Bootstrap methods: another look at the jackknife. **Annals of Statistics**, v.7, p.1-26, 1979.
- FERNANDES, L.M. **Inferência bayesiana em modelos discretos com fração de cura**. 2014. 62f. Dissertação (Mestrado) - Departamento de Estatística, Universidade de Brasília, 2014.
- HAMMERSLEY, J.M.; HANDSCOMB, D.C. **Monte Carlo methods**. London: Methuen. 1964. 178 p.
- JONHSON, N.L.; KOTZ, S. **Distributions in Statistics: continuous distributions**. John Wiley and Sons, New York, 1970.

KALBFLEISCH, J.D.; PRENTICE, R.L. **The Statistical Analysis of Failure Time Data**. New York : John Wiley & Sons, 1980.

LAWLESS, J.F. **Statistical models and methods for lifetime data**, New York: Wiley, 1982.

LOUZADA-NETO, F.; PERDONÁ, G.S.C.; CARRASCO, C.G. The Bi-log-logistic Model – A comparison study of some approximate confidence interval procedures. **Journal of Statistical Theory and Applications**, v.8, n.4, p.478-492, 2009.

MARUNBINI, E.; VALSECCHI, M.G. **Analysis survival data from clinical trials and observation studies**. England: John Wiley and Sons, 1995.

NAKANO, E.Y.; CARRASCO, C.G. Uma avaliação do Uso de um Modelo Contínuo na Análise de Dados Discretos de Sobrevida, **TEMA - Tend. Mat. Apl. Comput.**, v.7, n.1, p.91-100, 2006.

MELLO, M.P.; PETERNELLI, L.A. **Conhecendo o R: uma visão mais que estatística**. Viçosa:UFV, 2013.

R CORE TEAM **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: <http://www.R-project.org/> .

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, v.52, n.3, p.591-611, 1965.

WERNECK, A A.et al. Time to rehospitalization in patients with discharged on first generation antipsychotics, non-clozapine second generation antipsychotics, or clozapine. **Psychiatry Research** , v.188, n.3, p.315-319, 2011.