



The effect of the presence of censorship an application in lifetime data of electrical lamps

Tainara Nadingy da Silva Pereira¹ e Paulo José Pereira²

^{1,2}Universidade Federal do Vale do São Francisco (UNIVASF).

Email: tainaranadiny@hotmail.com, paulo.pereira@univasf.edu.br

Received: February 13th, 2017

Accepted: March 11th, 2017

Published: June 30th, 2017

Copyright ©2016 by authors and Institute of Technology Galileo of Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>

ABSTRACT

The Survival Analysis is the set of statistical techniques that evaluate the time until the occurrence of a certain event, even with the presence of elements that do not provide complete information about their respective times, called censored data. The objective of this work is to present the effects of the insertion of censored data in the reliability analysis of electric lamps using three non-parametric estimators, Kaplan-Meier, Nelson-Aalen and Life Table as comparative methods. The results showed that the difference between the survival function with the censored and uncensored data is more intense in the case of the presence of type I censorship. It was verified that the estimators tend to overestimate the survival with the censorship increase, independently of the Type of this, however, they maintain the same downward trend of survival over time. In the overestimation question, the estimator Nelson-Aalen showed ahead of the others.

Key Words: Censorship, Kaplan-Meier, Life Table, Nelson-Aalen.

O efeito da presença de censura: uma aplicação em dados de tempo de vida de lâmpadas elétricas

RESUMO

A Análise de Sobrevivência é o conjunto de técnicas estatísticas que avalia o tempo até a ocorrência de um determinado evento, mesmo com a presença de elementos que não forneçam a informação completa sobre seus respectivos tempos, denominados dados censurados. Este trabalho temo objetivo de apresentar os efeitos da inserção de dados censurados na análise de confiabilidade de lâmpadas elétricas, utilizando, como métodos comparativos, três estimadores não paramétricos, Kaplan-Meier, Nelson-Aalen e Tabela de Vida. Os resultados mostraram que a diferença entre a função de sobrevivência com os dados censurados e os sem censura é mais intensa no caso da presença de censura do tipo I. Foi verificado que os estimadores tendem a superestimar a sobrevida com o acréscimo de censura, independentemente do tipo desta, porém, eles mantêm a mesma tendência de sobrevivência decrescente ao longo do tempo. No quesito de superestimação, o estimador Nelson-Aalen se mostrou à frente dos demais.

Palavras-Chaves: Censura, Kaplan-Meier, Tabela de Vida, Nelson-Aalen.

I. INTRODUÇÃO

O avanço tecnológico tem permitido a evolução da indústria devido ao aprimoramento dos meios de produção e, a partir disso, a possibilidade de inclusão de diversas funções aos produtos fabricados. Em ambientes virtuais, por exemplo, a tecnologia proporciona uma proximidade das pessoas para com o mundo, permitindo o acesso rápido e fácil a qualquer tipo de conhecimento. Essa disponibilidade de informações estimula o consumidor a ter a necessidade de conhecer as especificações do produto requerido e as experiências de outros com o mesmo, provocando, assim, um crescente grau de exigência no mercado.

O acréscimo de várias funções ao produto provoca maiores dificuldades para a empresa na questão de garantir a execução de todas elas de forma satisfatória. Com base nisso e no alto grau de exigência do mercado, o nível competitivo das organizações se eleva, já que todas as companhias tendem a aprimorar sua produção visando conquistar a confiança do cliente. Para [1] adotam a definição de que "confiabilidade é a probabilidade de um item desempenhar satisfatoriamente a função requerida, sob condições de operação estabelecidas, por um período de tempo predeterminado". Logo, para possuir a garantia de um bom funcionamento, é necessária a existência de estimativas que relatem como o produto reage a tal situação e quanto tempo ele a



suporta. Estas informações podem ser obtidas a partir de testes de tempo de falha.

De acordo com [1], a análise de tempo de falha é um conjunto de técnicas estatísticas executado com o auxílio de informações a respeito da durabilidade do objeto em estudo. Os dados com tais informações podem ser obtidos em campo ou em testes de vida. Nesse último caso é simulado as mesmas condições observadas no campo no intuito de adquirir os tempos até a ocorrência da falha. O objetivo da análise de tempo de falha é estimar quantidades de interesse tais como: o tempo médio de falha, a fração esperada de falhas no período de garantia, etc.

Contudo, é possível ocorrer particularidades durante a realização dos testes, como por exemplo, a ocorrência de falhas devido a razões diferentes às analisadas ou por causa da ausência de falhas no período de tempo em que o teste foi programado. Essas situações provocam dados com informações incompletas, estes então são denominados dados censurados. Mesmo que incompletos esses dados devem ser considerados na análise. O objetivo deste trabalho foi analisar e identificar a influência da inclusão de dados censurados na análise de tempos de vida fazendo uso de tradicionais estimadores não paramétricos da Análise de Sobrevivência.

II. REFERENCIAL TEÓRICO

Segundo [2], “a Análise de Sobrevivência é o conjunto de técnicas estatísticas para analisar tempos de vida que envolve dados censurados (observações incompletas ou parciais sobre indivíduos que fizeram parte do experimento)”. Para [2] ainda explica que, “os dados de tempo de vida de indivíduos são aqueles medidos a partir do início do experimento até a perda ou falha destes”.

Ao longo dos anos esta metodologia passou a ter sua aplicação em diversas áreas do conhecimento, tais como:

- Medicina, onde o foco da análise é a trajetória do paciente. Nesses casos o tempo de falha pode ser entendido como o tempo até a morte do paciente, bem como até a cura ou recidiva de uma doença e a censura pode ser a morte por causas independentes à doença estudada. Um exemplo dessa aplicação pode ser encontrado no estudo de [3] sobre a sobrevivência de mulheres tratadas com câncer de mama no estado do Rio de Janeiro; e na pesquisa de [4] e [5] acerca dos fatores prognósticos de pacientes com câncer de próstata;
- Engenharia, na qual esse tipo de análise é mais conhecido como confiabilidade. O evento de interesse passa a ser a ocorrência de falha no produto. São estudos como esse que determinam o prazo de garantia aplicável a determinado produto, por exemplo. Uma aplicação similar pode ser conferida no trabalho de [5], neste a análise de confiabilidade é realizada em equipamentos elétrico-eletrônicos que possuem o tempo censurado. Outro estudo que pode ser citado é o de [6] que trata da aplicação da análise de sobrevivência a testes que avaliam a eficiência de preservativos de madeira;
- Ciências sociais, em que várias situações de interesse têm como resposta o tempo entre eventos. Como em relação à durabilidade de casamentos; o tempo falha nesse caso é o período em que o casamento resistiu ao divórcio, ou seja, a separação do casal é a falha. Essa situação foi abordada

por [7] em seu trabalho Análise de sobrevivência aplicada ao tempo de duração de casamentos. Já [2] focalizou seu estudo para os processos de adoção buscando descobrir se os perfis das crianças e adolescentes influenciam na duração do processo de adoção [8].

II.1. CENSURA

Para [8] ressaltam que a análise estatística deve considerar todos os resultados advindos dos testes de sobrevivência, independente da existência de censura ou não. De acordo com [1] existem três mecanismos de censura:

- Censura por tempo ou do tipo I: caracterizada pelo estabelecimento de um tempo limite para a realização do teste;
- Censura por falha ou do tipo II: é aquela que determina uma quantidade aceitável de falhas nos itens sob teste, sendo que depois de conseguido esse número de falhas, o teste é encerrado;
- Censura do tipo aleatório: esta ocorre devido a falhas por causas alheias àquelas estudadas ou quando os itens são excluídos do teste sem ter atingido a falha.

Esses mecanismos estão ilustrados na Figura 1:

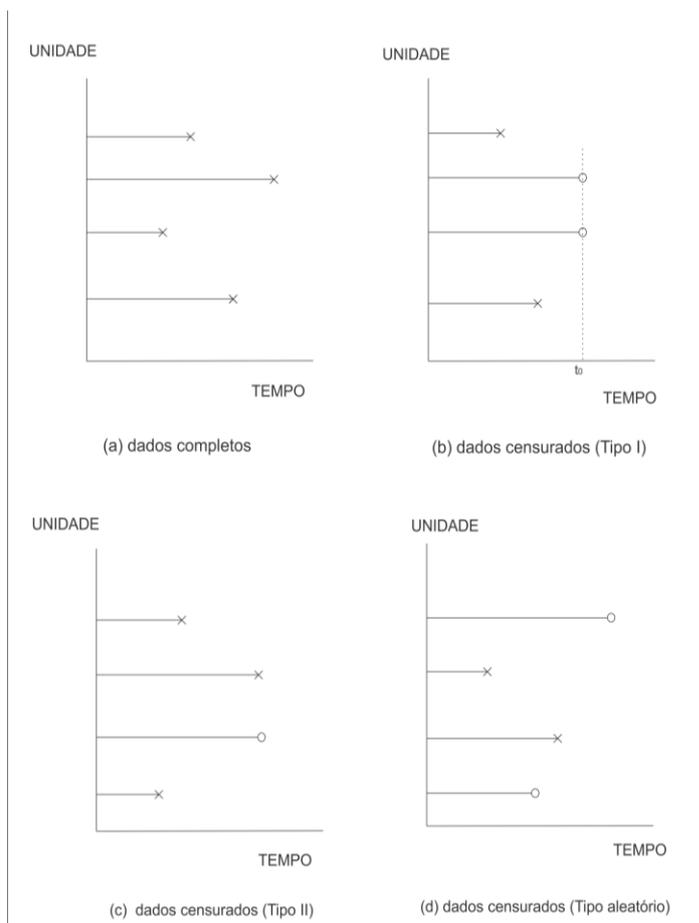


Figura 1: Tipos de dados de confiabilidade (X: falha; O: censura).
Fonte: Adaptado de [1].

II.2. FUNÇÃO SOBREVIVÊNCIA

De acordo com [8], a função de sobrevivência é definida como sendo a probabilidade de uma observação não falhar até certo tempo t [...]. Em termos probabilísticos isso é descrito como:

$$S(t) = P(T \geq t) \quad (1)$$

III. METODOLOGIA

III.1. INFERÊNCIA NÃO-PARAMÉTRICA

As funções de confiabilidade e de taxa de falha são facilmente estimadas quando não há presença de dados censurados. Por outro lado, com a existência dessas informações incompletas é exigida a utilização de técnicas especializadas para que seja possível realizar a análise. Essa alteração ocorre devido ao desconhecimento, provocado pelas censuras, referente à frequência exata associada a alguns intervalos. As técnicas que analisam os dados de tempo de falha na presença de censura e não necessitam especificar uma distribuição para a variável *tempo até a falha* são conhecidas como estimadores não-paramétricos [1]. Os estimadores não-paramétricos abordados neste trabalho foram: Tabela de vida, Kaplan-Meier e Nelson-Aalen.

III.2. ESTIMADOR TABELA DE VIDA

De acordo com [1], “a tabela de vida ou método atuarial é uma das mais antigas técnicas estatísticas utilizadas para estimar características associadas à distribuição dos tempos de falha”. Há registros de Tabelas de vida sendo utilizadas desde o século XVII, inicialmente por companhias de seguro.

Para [1] explicam que para construir uma tabela de vida é necessário sectionar o eixo do tempo em intervalos. Supondo que esse número seja $k+1$ intervalos definidos pelos pontos de corte, t_1, t_2, \dots, t_k e $t_0 = 0$. Para cada um dos intervalos vamos estimar a seguinte probabilidade:

$$q_i = P(T \in [t_{i-1}, t_i] / T \geq t_{i-1}) \quad (2)$$

isto é, q_i é a probabilidade de um item falhar no intervalo $[t_{i-1}, t_i]$ sabendo-se que ele sobreviveu até t_{i-1} . A partir destes valores obtemos a função de confiabilidade. [...]

A função de confiabilidade pode ser entendida como um percentual de certeza de que um item não falhará até o tempo t_i , $i = 1, \dots, k$. Isto é dado em termos de q 's como:

$$S(t_i) = (1 - q_1) \dots (1 - q_i) \quad (3)$$

III.3. ESTIMADOR KAPLAN-MEIER

Com base em [8], Kaplan e Meier propuseram a utilização do estimador não paramétrico que carrega os seus nomes em 1958 a partir da adaptação da função de sobrevivência empírica. Este estimador também é chamado de limite-produto e sua versão na ausência de censuras, é definida como:

$$\widehat{S}(t) = \frac{\text{no.de observações que não falharam até o tempo } t}{\text{no.total de observações no estudo}} \quad (4)$$

" $\widehat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $1/n$, em que n é o tamanho da amostra. Se existirem empates em certo tempo t , o tamanho do degrau fica multiplicado pelo número de empates" [8].

Ainda com base em [8] é entendido que a metodologia de construção do estimador de Kaplan-Meier consiste em utilizar o número de falhas distintas como a quantidade de intervalos de tempo. Os limites dos intervalos de tempo são os tempos de falha da amostra. Para os casos em que as censuras e falhas ocorrem simultaneamente, deve-se considerar que a censura ocorre imediatamente depois da falha, ou seja, o tempo de falha será menor que o tempo de censura. Em situações como esta situação é dito que há censuras e falhas empatadas.

O estimador de Kaplan-Meier de $S(t)$ é definido como:

$$\widehat{S}(t) = [(n_1 - d_1)/n_1][(n_2 - d_2)/n_2] \dots [(n_{t_0} - d_{t_0})/n_{t_0}] \quad (5)$$

Onde t_0 é o maior tempo de falha menor que t .

III.4. ESTIMADOR NELSON-AALEN

De acordo com [8], o estimador Nelson-Aalen foi proposto a princípio por além em 1972 e depois aprimorado por Aalen em 1978. Este tem como base a função de sobrevivência expressa por:

$$S(t) = \exp\{-\Lambda(t)\}, \quad (6)$$

em que $\Lambda(t)$ é a função de risco acumulado.

E possui a seguinte forma:

$$\widehat{\Lambda}(t) = \sum_{j:t_j < t} \left(\frac{d_j}{n_j}\right), \quad (7)$$

III.5. PROCESSO METODOLÓGICO

Inicialmente, foram realizadas pesquisas acerca da análise de sobrevivência com aplicações em várias áreas do conhecimento. Através destas foi encontrado o estudo de [9] que trata da determinação da confiabilidade de lâmpadas elétricas por meio da função de distribuição de Weibull. Os autores utilizaram dados colhidos em um dos ensaios realizados mensalmente por uma fábrica de lâmpadas. Estes mesmos dados foram disponibilizados no trabalho e escolhidos para servir de base para a presente pesquisa. No intuito de avaliar o comportamento dos três estimadores citados perante a censura, foram utilizados dois tipos desta, a censura do tipo 1 e a aleatória [10]. Cada tipo de censura foi submetido a três situações, a cada situação a porcentagem de dados censurados era dobrada, os percentuais utilizados foram 5%, 10% e 20% do maior tempo falha do teste. Para a censura aleatória foi realizado um sorteio por meio do software Microsoft Office Excel®, versão 2010, com o propósito de conhecer os elementos que seriam censurados. O quantitativo de elementos sorteados foi equivalente ao número de ocorrências englobadas pelo tempo de falha censurado.

Após o conhecimento dos dados que seriam censurados, foi realizada a análise dos estimadores Kaplan Meier e Nelson Aalen por meio do software RGUI®, versão para Microsoft Windows. Já a estimativa da Tabela de Vida foi efetuada com auxílio do software BioEstat 5.0®.

IV. RESULTADOS E DISCUSSÕES

A Tabela 1 apresenta os resultados obtidos no estudo de Dias e Leite na ausência de censura.

Tabela 1: Estimação da sobrevivência na ausência de censura.

Tempo falha	Nº Risco	Ocorrências	Estimativa S(t)	IC - 95%
1394	20	2	0,9166	0,7838 a 0,9819
1451	18	2	0,8186	0,6563 a 0,9286
1470	16	1	0,7696	0,5989 a 0,8959
1501	15	1	0,7205	0,5444 a 0,8604
1550	14	1	0,6715	0,4921 a 0,8226
1591	13	2	0,5735	0,3935 a 0,7413
1706	11	1	0,5245	0,3469 a 0,6980
1750	10	3	0,3774	0,2170 a 0,5580
1773	7	1	0,3284	0,1773 a 0,5078
1797	6	2	0,2303	0,1040 a 0,4010
1888	4	2	0,1323	0,0412 a 0,2826
1980	2	2	0,0343	0,0025 a 0,1391

Fonte: Adaptado de [9].

As tabelas de resultados de todos os estimadores fornecem as mesmas análises. Na coluna referente ao tempo de falha é encontrado o sinal de soma (+) em alguns períodos, este é o indicador de que houve pelo menos uma censura naquele tempo. Na coluna de “Nº Risco” é apresentado o quantitativo de elementos vivos no início daquele intervalo. As ocorrências são os eventos de interesse, ou seja, as falhas. A taxa de risco indica as chances daqueles que estão vivos de falharem antes do final do período. O intervalo de confiança (IC) para estes dados possui 95% de confiança.

IV.1. CENSURA TIPO 1

As censuras deste tipo estão sempre situadas nas últimas observações. Os dados mostram que o percentual de sobrevivência é decrescente ao longo do tempo, como era esperado. Verifica-se

que a probabilidade de sobrevivência é conservada quando todos os dados do próximo intervalo são censurados, além disso, nesta mesma situação, a taxa de risco é nula.

IV.1.1. ANÁLISE COM 5% DE TEMPO CENSURADO

A Tabela 2 está relacionada à inserção de 5% de censura do tipo I, o que equivale a quatro elementos do banco de dados. Os resultados dos estimadores Kaplan-Meier e Tabela de vida expressam valores idênticos quanto à sobrevivência, porém os intervalos de confiança são um pouco divergentes. Já o estimador Nelson-Aalen apresenta valores similares, referente à sobrevivência, mesmo que um pouco maiores, cerca de 0,2% no início e 2% nas últimas observações.

Tabela 2: Análise de sobrevivência por meio dos três estimadores não paramétricos com 5% do tempo com censura do tipo I.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394	20	2	0,9	0,7777 a 1	0,902	0,7828 a 1	0,9	0,7685 a 1
1451	18	2	0,8	0,6426 a 0,996	0,805	0,6505 a 0,996	0,8	0,6548 a 0,9452
1470	16	1	0,75	0,5823 a 0,966	0,756	0,5914 a 0,967	0,75	0,6314 a 0,8686
1501	15	1	0,7	0,5254 a 0,933	0,707	0,5355 a 0,934	0,7	0,5738 a 0,8262
1550	14	1	0,65	0,4712 a 0,897	0,659	0,4823 a 0,899	0,65	0,5151 a 0,7849
1591	13	2	0,55	0,37 a 0,818	0,561	0,3827 a 0,823	0,55	0,3539 a 0,7461
1706	11	1	0,5	0,3226 a 0,775	0,512	0,3359 a 0,781	0,5	0,3301 a 0,6699
1750	10	3	0,35	0,1926 a 0,636	0,366	0,2072 a 0,647	0,35	0,066 a 0,634
1773	7	1	0,3	0,1536 a 0,586	0,317	0,1683 a 0,599	0,3	0,0408 a 0,5592
1797	6	2	0,2	0,0832 a 0,481	0,22	0,0974 a 0,496	0,2	0 a 0,5772
1888+	4	0	0,2	0,0832 a 0,482	0,22	0,0974 a 0,497	0,2	0,2 a 0,2
1980+	2	0	0,2	0,0832 a 0,483	0,22	0,0974 a 0,498	0,2	0,2 a 0,2

Fonte: Autores, (2016).

Considerando as estimativas de sobrevivência com dados não censurados fornecidos na Tabela 1 e as comparando com os resultados obtidos pelos três estimadores não paramétricos após a inserção de 5% de censura na Tabela 2, observa-se que as estimativas de sobrevivência dos três estimadores demonstraram desempenho semelhante nesta situação. Porém, nos tempos falha

1888 e 1980, nos quais as censuras são inseridas, as estimativas se conservam e a efetividade da análise é afetada. Para o estimador Kaplan-Meier e Tabela de Vida, as diferenças entre os resultados destas duas tabelas são de 6,77% e 16,57%, para os tempos 1888 e 1980 respectivamente, e 8,77% e 18,57% para o estimador Nelson-Aalen.

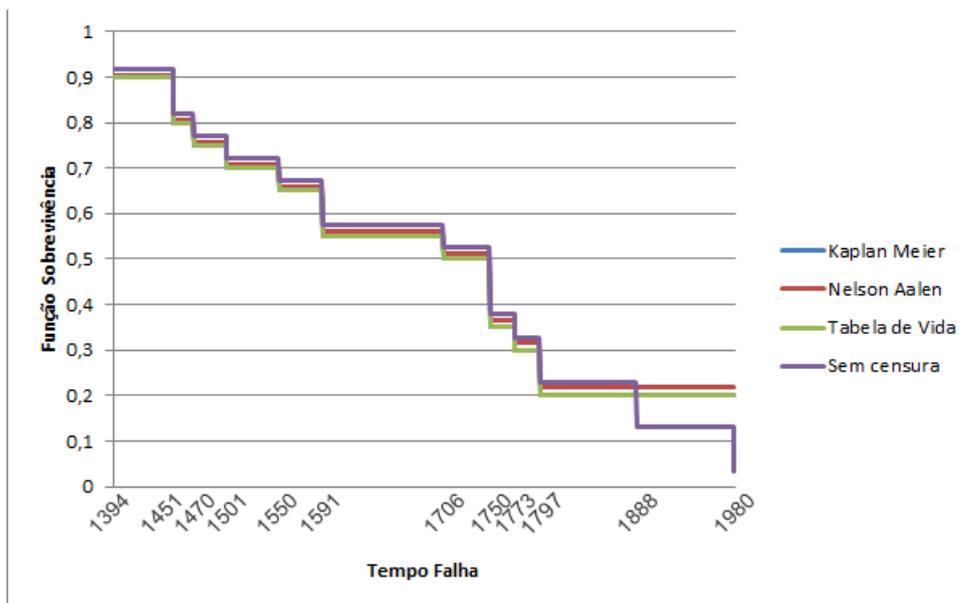


Figura 2: Funções de sobrevivência com 5% de censura do tipo I.
Fonte: Autores, (2016).

IV.1.2. ANÁLISE COM 10% DE TEMPO CENSURADO

A Tabela 3 se refere às análises com os dados submetidos a 10% do seu tempo censurado. Este percentual equivale a seis elementos censurados. É perceptível que apesar de dobrada a porcentagem de censuras inseridas, os valores iniciais das análises

se mantém os mesmos do cenário anterior. Isto se deve a característica da censura do tipo 1 de limitar o tempo da análise, afetando assim as últimas observações. É observado também que os intervalos de confiança se alteram de forma sutil nas primeiras observações que não são afetadas pela censura e de modo mais intenso naquelas afetadas.

Tabela 3: Análise de sobrevida por meio dos três estimadores não paramétricos com 10% do tempo com censura do tipo I.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de Vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394	20	2	0,9	0,778 a 1	0,902	0,783 a 1	0,9	0,7685 a 1
1451	18	2	0,8	0,643 a 0,996	0,805	0,65 a 0,996	0,8	0,6548 a 0,9452
1470	16	1	0,75	0,582 a 0,966	0,756	0,591 a 0,967	0,75	0,6314 a 0,8686
1501	15	1	0,7	0,525 a 0,933	0,707	0,536 a 0,934	0,7	0,5738 a 0,8262
1550	14	1	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,5151 a 0,7849
1591	13	2	0,55	0,37 a 0,818	0,561	0,383 a 0,823	0,55	0,3539 a 0,7461
1706	11	1	0,5	0,323 a 0,775	0,512	0,336 a 0,781	0,5	0,3301 a 0,6699
1750	10	3	0,35	0,193 a 0,636	0,366	0,207 a 0,647	0,35	0,0606 a 0,634
1773	7	1	0,3	0,154 a 0,586	0,317	0,168 a 0,599	0,3	0,0408 a 0,5592
1797+	6	0	0,3	0,154 a 0,586	0,317	0,168 a 0,599	0,3	0,3 a 0,3
1888+	4	0	0,3	0,154 a 0,586	0,317	0,168 a 0,599	0,3	0,3 a 0,3
1980+	2	0	0,3	0,154 a 0,586	0,317	0,168 a 0,599	0,3	0,3 a 0,3

Fonte: Autores, (2016).

As divergências entre os resultados dos dados censurados e os que não são foram intensificadas devido ao fato de que os 10% de censura alcançaram um novo tempo falha, conservando assim, a sobrevida obtida no tempo anterior a este. Com isso, as diferenças entre os resultados de Kaplan-Meier, Tabela de vida e os originais

sem censura passaram a ser 6,97%; 16,77%; 26,57% para os respectivos tempos falha 1797, 1888, 1980. Já referente ao Nelson-Aalen, para os mesmos tempos, a diferença é de 8,67%; 18,47%; 28,27%.

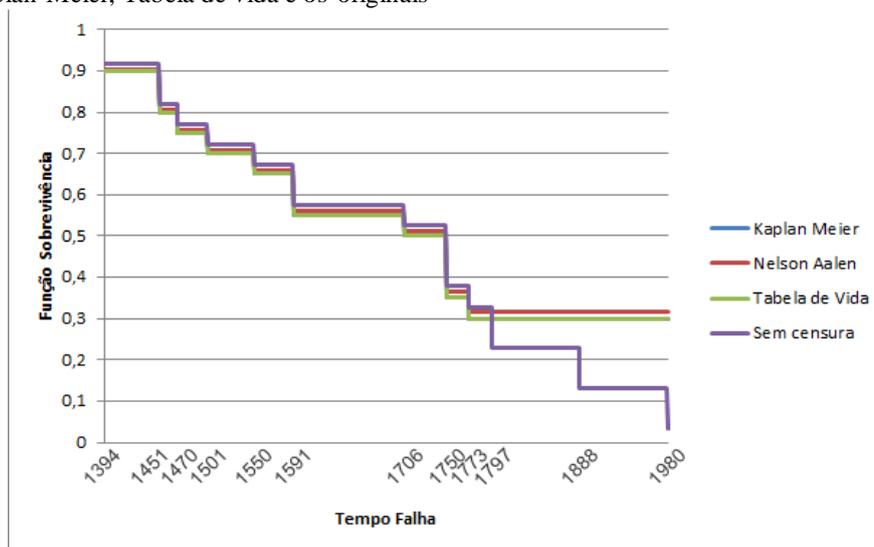


Figura 3 - Funções de sobrevivência com 10% de censura do tipo I
Fonte: Autores, (2016).

IV.1.3. ANÁLISE COM 20% DE TEMPO CENSURADO

Redobrando a porcentagem, o tempo censurado chega a 20%, o equivalente ao alcance de 13 elementos. Essa porcentagem de censura afeta 58,33% dos tempos falha e 65% dos elementos analisados. Estes valores indicam que os dados em risco nos tempos afetados terão suas sobrevidas superestimadas, pois suas estimativas serão baseadas na sobrevida do tempo falha 1550. As discrepâncias das estimativas se iniciam no tempo falha 1591 com uma diferença de 7,65% comparando o estimador Kaplan-Meier e

Tabela de Vida com os dados sem censura, e 8,55% para o estimador Nelson-Aalen, até o tempo 1980 com uma diferença de 61,57% para Kaplan-Meier e Tabela de Vida e 62,47% para Nelson-Aalen.

Todas essas informações podem ser conferidas na Tabela 6. A diferença média entre os dados originais e os resultados obtidos nessa etapa foi: 18,77% para os estimadores Kaplan-Meier e Tabela de vida, 19,53% para Nelson-Aalen.

Tabela 4: Análise de sobrevida por meio dos três estimadores não paramétricos com 20% do tempo com censura do tipo I.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de Vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394	20	2	0,9	0,778 a 1	0,902	0,783 a 1	0,9	0,7685 a 1
1451	18	2	0,8	0,643 a 0,996	0,805	0,65 a 0,996	0,8	0,6548 a 0,9452
1470	16	1	0,75	0,582 a 0,966	0,756	0,591 a 0,967	0,75	0,6314 a 0,8686
1501	15	1	0,7	0,525 a 0,933	0,707	0,536 a 0,934	0,7	0,5738 a 0,8262
1550	14	1	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,5151 a 0,7849
1591+	13	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1706+	11	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1750+	10	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1773+	7	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1797+	6	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1888+	4	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65
1980+	2	0	0,65	0,471 a 0,897	0,659	0,482 a 0,899	0,65	0,65 a 0,65

Fonte: Autores, (2016).

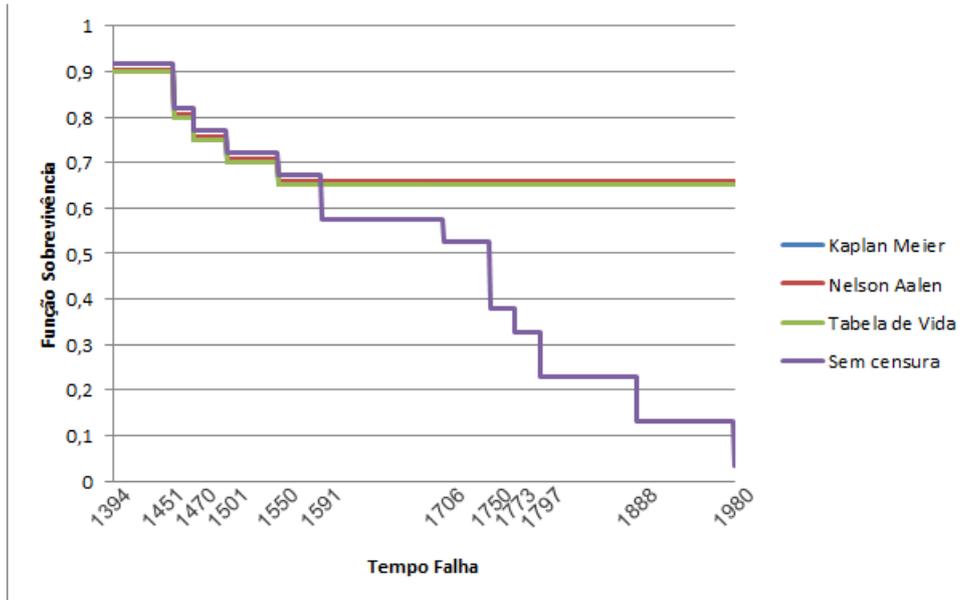


Figura 4 - Funções de sobrevivência com 20% de censura do tipo I
Fonte: Autores, (2016).

IV.2. CENSURA ALEATÓRIA

IV.2.1. ANÁLISE COM 5% DE TEMPO CENSURADO

A Tabela 9 apresenta as análises de sobrevivência com 5% do tempo censurado de modo aleatório. Os elementos sorteados e censurados foram: 5, 7, 10 e 20; nos tempos de falha 1470, 1550 e 1706, respectivamente. Coincidentemente, os elementos censurados eram as únicas ocorrências de seus respectivos tempos falha, assim, as sobrevidas desses tempos se tornaram iguais às

sobrevidas dos tempos que os antecediam. O estimador Tabela de Vida apresenta a sobrevida dos dados que receberam censura com 100% de certeza, pois a amplitude do intervalo de confiança é nula nesses tempos.

As sobrevidas dos estimadores não paramétricos são mais otimistas do que as originais. Essa superestimação dos elementos é mais intensa no tempo 1706 por alcançar 10,73% de discrepância, para o estimador Kaplan-Meier e Tabela de Vida; e 11,7% para o estimador Nelson-Aalen.

Tabela 5: Análise de sobrevida por meio dos estimadores não paramétricos com 5% do tempo censurado de modo aleatório.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de Vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394	20	2	0,9	0,77767 a 1	0,9025	0,7828 a 1	0,9	0,7685 a 1
1451	18	2	0,8	0,64257 a 0,996	0,8049	0,6505 a 0,996	0,8	0,6548 a 0,9452
1470+	16	0	0,8	0,64257 a 0,996	0,8049	0,6505 a 0,996	0,8	0,8 a 0,8
1501	15	1	0,7467	0,57715 a 0,966	0,753	0,5865 a 0,967	0,7467	0,6204 a 0,8729
1550+	14	0	0,7467	0,57715 a 0,966	0,753	0,5865 a 0,967	0,7467	0,7467 a 0,7467
1591	13	2	0,6318	0,4468 a 0,893	0,6415	0,4591 a 0,896	0,6318	0,4357 a 0,8279
1706+	11	0	0,6318	0,4468 a 0,893	0,6415	0,4591 a 0,896	0,6318	0,6318 a 0,6318
1750	10	3	0,4423	0,25939 a 0,754	0,4584	0,2759 a 0,762	0,4423	0,1582 a 0,7263
1773	7	1	0,3791	0,20529 a 0,7	0,3974	0,2225 a 0,71	0,3791	0,1198 a 0,6383
1797	6	2	0,2527	0,10971 a 0,582	0,2754	0,1272 a 0,596	0,2527	0 a 0,6164
1888	4	2	0,1269	0,03488 a 0,458	0,1537	0,0499 a 0,473	0,1264	0 a 0,6299
1980+	2	1	0,0632	0,00953 a 0,419	0,0932	0,021 a 0,414	0,0421	0 a 0,7965

Fonte: Autores, (2016).

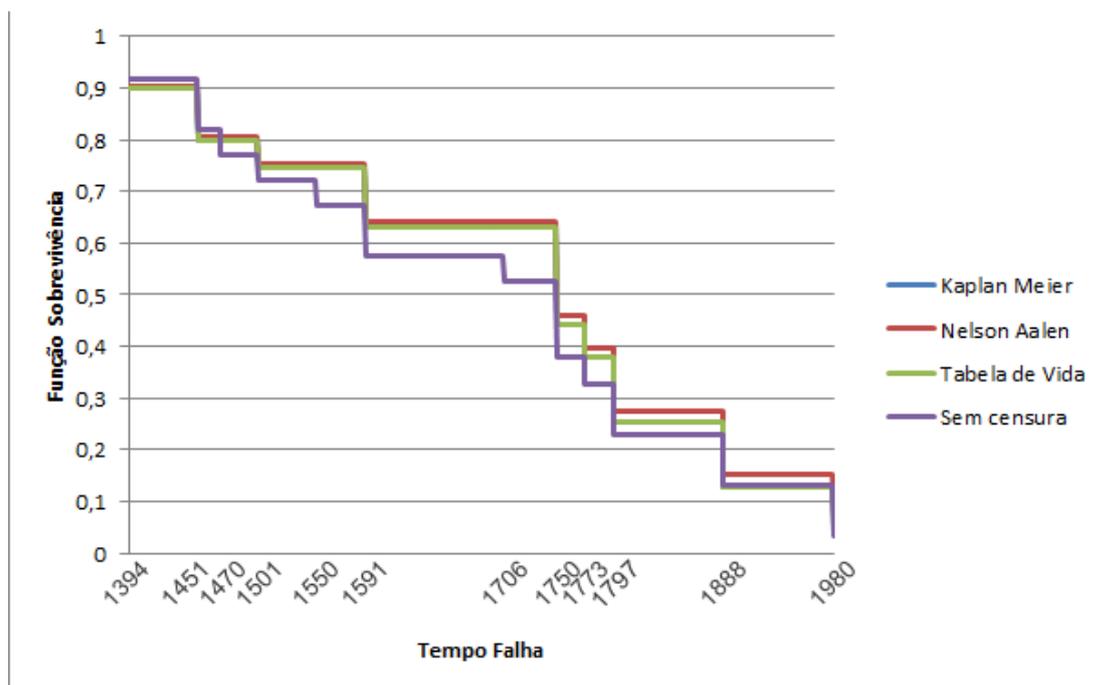


Figura 5: Funções de sobrevivência com 5% de censura aleatória.
 Fonte: Autores, (2016).

IV.2.2. ANÁLISE COM 10% DE TEMPO CENSURADO

Os resultados das análises de sobrevivência referentes aos 10% de censura do tipo aleatório estão expressos na Tabela 12. A partir do tempo de falha 1750 as estimativas de Kaplan-Meier e

Tabela de Vida passam a ser distintas, e a divergência entre as análises com e sem censura se intensificam. As diferenças das estimativas no tempo 1706 se mantêm, porém a diferença no tempo 1750 chega a 12,8% no estimador Kaplan-Meier, 14,16% para o Nelson-Aalen e 12,16% para a Tabela de Vida.

Tabela 4: Análise de sobrevida por meio dos estimadores não paramétricos com 10% do tempo censurado de modo aleatório.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394	20	2	0,9	0,7777 a 1	0,902	0,7828 a 1	0,9	0,7685 a 1
1451	18	2	0,8	0,64257 a 0,996	0,805	0,6505 a 0,996	0,8	0,6548 a 0,9452
1470+	16	0	0,8	0,64257 a 0,996	0,805	0,6505 a 0,996	0,8	0,8 a 0,8
1501	15	1	0,7467	0,57715 a 0,966	0,753	0,5865 a 0,967	0,7467	0,6204 a 0,8729
1550+	14	0	0,7467	0,57715 a 0,966	0,753	0,5865 a 0,967	0,7467	0,7467 a 0,7467
1591	13	2	0,6318	0,4468 a 0,893	0,642	0,4591 a 0,896	0,6318	0,4357 a 0,8279
1706+	11	0	0,6318	0,4468 a 0,893	0,642	0,4591 a 0,896	0,6318	0,6318 a 0,6318
1750+	10	2	0,5054	0,3175 a 0,805	0,519	0,333 a 0,81	0,4988	0,2396 a 0,758
1773	7	1	0,4332	0,2488 a 0,754	0,45	0,2662 a 0,762	0,4275	0,1683 a 0,6868
1797	6	2	0,2888	0,1308 a 0,638	0,312	0,15 a 0,649	0,285	0 a 0,6622
1888+	4	1	0,2166	0,0818 a 0,573	0,243	0,1007 a 0,587	0,2036	0 a 0,6769
1980+	2	1	0,1083	0,0199 a 0,589	0,147	0,0395 a 0,551	0,0679	0 a 0,8223

Fonte: Autores, (2016).

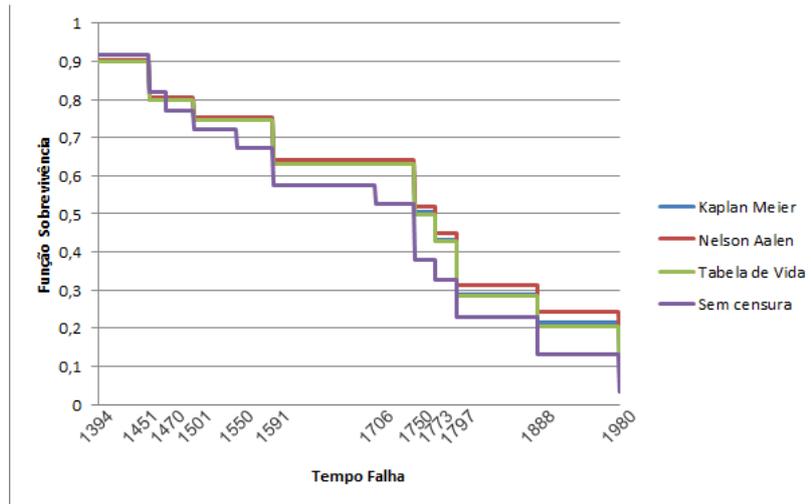


Figura 6: Funções de sobrevivência com 10% de censura aleatória.
Fonte: Autores, (2016).

IV.2.3. ANÁLISE COM 20% DE TEMPO CENSURADO

A análise que representa 20% dos elementos com mecanismo de censura aleatório é a última situação a ser analisada, os resultados desta se encontram na Tabela 15. Neste caso os efeitos são de alta proporção o que provoca uma superestimação da confiabilidade desse produto. Todas as estimativas são

prejudicadas. As médias de erro são: 24,13% para Kaplan-Meier, 25,19% para Nelson-Aalen e 23,27% para a Tabela de Vida. As estimativas mais errôneas se encontram no tempo falha 1980 com as diferenças de 35,32%, 38,47% e 32,31%, seguindo a mesma ordem.

Tabela 7: Análise de sobrevida por meio dos três estimadores não paramétricos com 20% do tempo censurado de modo aleatório.

Tempo falha	Nº Risco	Ocorrências	Kaplan Meier		Nelson Aalen		Tabela de vida	
			S(t)	IC - 95%	S(t)	IC - 95%	S(t)	IC - 95%
1394+	20	1	0,95	0.859 a 1	0,951	0.862 a 1	0,9487	0.8508 a 1
1451+	18	0	0,95	0.859 a 1	0,951	0.862 a 1	0,9487	0.9487 a 0.9487
1470+	16	0	0,95	0.859 a 1	0,951	0.862 a 1	0,9487	0.9487 a 0.9487
1501+	15	0	0,95	0.859 a 1	0,951	0.862 a 1	0,9487	0.9487 a 0.9487
1550+	14	0	0,95	0.859 a 1	0,951	0.862 a 1	0,9487	0.9487 a 0.9487
1591	13	2	0,8038	0.624 a 1	0,810	0.636 a 1	0,8028	0.6066 a 0.9989
1706+	11	0	0,8038	0.624 a 1	0,810	0.636 a 1	0,8028	0.8028 a 0.8028
1750+	10	1	0,7234	0.522 a 1	0,733	0.537 a 1	0,7136	0.5082 a 0.9189
1773	7	1	0,6201	0.397 a 0.968	0,636	0.418 a 0.967	0,6116	0.3524 a 0.8709
1797	6	1	0,5167	0.292 a 0.915	0,538	0.316 a 0.916	0,5004	0.1781 a 0.8228
1888+	4	1	0,3876	0.173 a 0.866	0,419	0.203 a 0.863	0,3574	0 a 0.8307
1980+	2	0	0,3875	0.173 a 0.866	0,419	0.203 a 0.863	0,3574	0.3574 a 0.3574

Fonte: Autores, (2016).

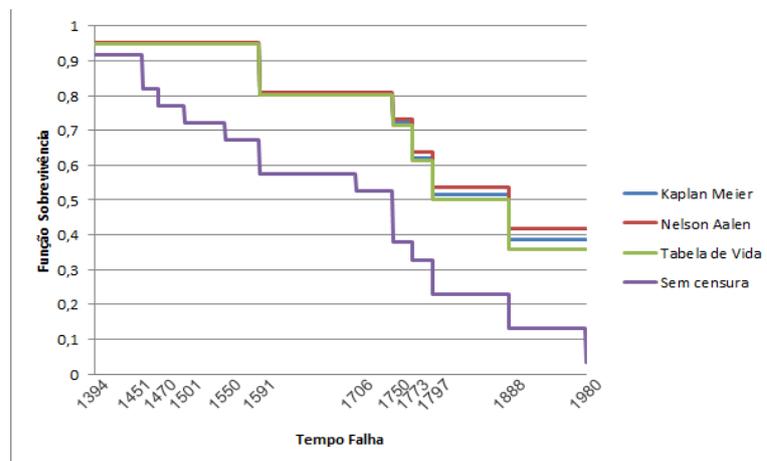


Figura 7: Funções de sobrevivência com 20% de censura aleatória.
Fonte: Autores, (2016).

V. CONSIDERAÇÕES FINAIS

Os estimadores não paramétricos possuem comportamentos semelhantes quando submetidos à censura, demonstrando sobrevidas decrescentes ao longo do tempo e os dados censurados estimam sobrevidas mais elevadas que as originais. A superestimação ocorre nos dois tipos de censura, porém ela ocorre de modo mais intenso na censura do tipo I, mesmo que seja mais facilmente percebida nos casos de censura do tipo aleatória. Dos três estimadores, Nelson-Aalen foi o que relatou as maiores estimativas em todas as situações dos dois tipos de censura trabalhados. Os resultados mais divergentes entre Nelson-Aalen e os outros estimadores se encontram no momento em que são inseridos 10% de censura do tipo aleatório, sendo 3,87% em relação ao Kaplan-Meier e 7,91% com a Tabela de Vida.

A presença de censura do tipo I afeta apenas as sobrevidas dos dados que receberam a censura, ou seja, as últimas estimativas. Observando o último tempo falha, 1980, censurado nos três casos apresentados de censura de tipo I, é visto o modo como a censura influencia a estimativa. Com base no estudo original, na ausência de censuras a sobrevida neste tempo foi de 3,43%. Para Kaplan-Meier e Tabela de vida, as estimativas obtidas foram 2%, 3% e 65%, para os respectivos casos com 5%, 10% e 20% de censura. Para Nelson-Aalen as sobrevidas foram: 2,2%, 3,17% e 65,9%, respeitando a mesma ordem. A partir de um simples comparativo é perceptível que a reação das estimativas não respeita a proporção com que as censuras são inseridas, ou seja, duplicando-se.

Já as censuras aleatórias influenciam até mesmo os dados que não foram censurados, por se encontrarem em um tempo de falha maior que o do elemento censurado. Um exemplo desta situação é o tempo falha 1797, que mesmo não obtendo nenhuma censura, sua sobrevida cresce 3,61% do primeiro caso para o segundo e 22,79% do segundo para o terceiro, considerando as estimativas de Kaplan-Meier; 3,66% e 22,6% segundo Nelson-Aalen e de acordo com a Tabela de Vida estes aumentos foram de 3,23% e 21,54%.

Por outro lado, os tempos falha censurados desde a inserção de 5% de informações incompletas, como o tempo 1470, não sofreram alterações em sua sobrevida do primeiro para o segundo caso, somente do segundo para o terceiro. Com o olhar voltado para outro tempo falha que também recebeu censura aleatória desde o início, como o 1550, a alteração sofrida no segundo momento tem maior proporção do que aquela sofrida pelo tempo 1470. Com isso, pode-se afirmar que, se tratando deste tipo de censura, não é apenas a quantidade de dados censurados que afetam a confiabilidade da análise, mas também a posição destes na linha do tempo do teste.

VI. REFERÊNCIAS

- [1] Freitas, Marta Afonso; Colosimo, Enrico Antônio. **Confiabilidade: análise de tempo de falha e testes de vida acelerados**. Belo Horizonte: Fundação Christiano Ottoni, Escola de engenharia da UFMG, 1997.
- [2] Pereira, Paulo José. **Estudo da viabilidade da aplicação da análise de sobrevivência em dados entomológicos**. Tese (Mestrado Agronomia – Estatística e Experimentação Agropecuária). Lavras: UFLA, 2002.
- [3] Brito, Claudia; Portela, Margareth; Vasconcelos, Maurício. **Sobrevida de mulheres tratadas por câncer de mama no estado do Rio de Janeiro**. Revista Saúde Pública, Rio de Janeiro, 2009.
- [4] Migowski, Arn; Silva, Gulnar. **Sobrevida e fatores prognósticos de pacientes com câncer de próstata clinicamente localizado**. Revista Saúde Pública, Rio de Janeiro, 2010.
- [5] Silva, Josiane. et al. **Estudo da confiabilidade de equipamento elétrico-eletrônico com tempo censurado**. Encontro Nacional de Engenharia Produção. Anais. Fortaleza, Ceará, 2015.
- [6] Matos, Carlos. **Modelagem não paramétrica de análise de sobrevivência aplicada aos testes de estaca usados na avaliação da eficiência de preservativos de madeira**. Encontro Brasileiro em Madeiras e em Estruturas de Madeira. Anais. Curitiba, Paraná, 2016.
- [7] Simões, Eivaldo. **Análise de sobrevivência aplicada a casamentos**. Trabalho de Conclusão de Curso. Campina Grande: Universidade Estadual da Paraíba, 2014.
- [8] Colosimo, Enrico Antônio; Giolo, Suely Ruiz. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006.
- [9] Dias, José & Leite, Maria. **Uma aplicação de confiabilidade na indústria de lâmpadas elétricas**. Encontro Nacional de Engenharia Produção. Anais. 1998.
- [10] Pereira, Paulo José. **Adoção realidades e desafios para um Brasil do século XXI**. Tese (Doutorado em Demografia). Campinas: Universidade Estadual de Campinas, 2012.