



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

Erivaldo dos Santos Simões

# Análise de sobrevivência aplicada ao tempo de duração de casamentos

Campina Grande  
Dezembro de 2014

Erivaldo dos Santos Simões

# **Análise de sobrevivência aplicada ao tempo de duração de casamentos**

Trabalho Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Dr. Tiago Almeida de Oliveira

Campina Grande  
Dezembro de 2014

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S589a Simões, Erivaldo dos Santos.

Análise de sobrevivência aplicada ao tempo de duração de casamentos [manuscrito] / Erivaldo dos Santos Simões. - 2014. 43 p. : il.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Análise de sobrevivência. 2. Métodos paramétricos. 3. Kaplan – Meier. 4. Divórcio. I. Título.

21. ed. CDD 519.544

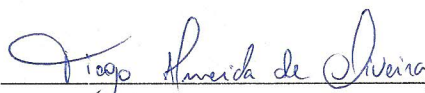
Erivaldo dos Santos Simões

## Análise de sobrevivência aplicada ao tempo de duração de casamentos

Trabalho Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 18 / 12 / 14

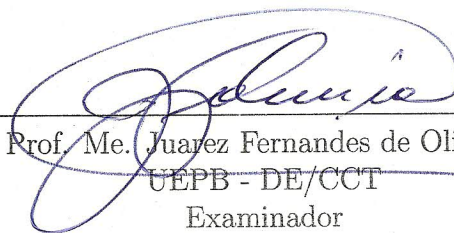
### Banca Examinadora:



Prof. Dr. Tiago Almeida de Oliveira  
UEPB - DE/CCT  
Orientador



Prof. Dra. Ana Patricia Bastos Peixoto  
UEPB - DE/CCT  
Examinador



Prof. Me. Juarez Fernandes de Oliveira  
UEPB - DE/CCT  
Examinador

Dedico esse trabalho a minha família.

# Agradecimentos

Agradeço a Deus em primeiro lugar, presente em minha vida em todos os desafios, me dando forças e me rodeando de pessoas do bem que sempre me ajudaram a superar um a um.

Agradeço a todos os mestres que tiveram paciência e me acrescentaram algo em toda a jornada na faculdade, em especial ao Doutor Tiago Almeida de Oliveira, professor e amigo que aceitou me orientar neste trabalho, e sua esposa, Prof. Dra. Ana Patricia Bastos Peixoto, que quando pensei em desistir, foi a injeção de ânimo e força.

Agradeço aos amigos que me ajudaram a superar percalços durante todo o tempo que estive em busca de minha formação dos quais não posso citar nomes, pois provavelmente iria pecar por esquecimento.

Deus é bom.

# Resumo

Análise de Sobrevivência é a expressão utilizada para designar a análise estatística de dados quando a variável em estudo representa o tempo desde um instante inicial bem definido até à ocorrência de determinado acontecimento de interesse. No presente trabalho procurou-se utilizar ferramentas estatísticas em especial a análise de sobrevivência na busca de uma distribuição paramétrica que modelasse o tempo  $t$  até o divórcio sabendo que a função de sobrevivência não paramétrica de Kaplan-Meier se adequou aos dados conforme estudo referenciado. O casamento e divórcio de forma geral e especialmente no Brasil foram explorados, além de todo um levantamento bibliográfico dos métodos não-paramétricos e paramétricos com suas funções de distribuição de probabilidade. Utilizamos um banco de dados coletados no 1<sup>o</sup> cartório do registro civil Severino Cavalcante na cidade de Campina Grande em 25/11/2001. Os resultados encontrados sugerem que o estimador de Kaplan-Meier foi o único que modelou o tempo até o divórcio, nenhuma das distribuições estudadas obteve resultados satisfatórios.

**Palavras - Chave:** Divórcio; Kaplan - Meier; Métodos Paramétricos.

# Abstract

In this paper , we use statistical tools , in particular, the survival analysis in search of a parametric distribution to model the time  $t$  in the divorce knowing that the Kaplan-Meier survival function if adapted the data of the referenced study. We use the database collected in the 1 office of civil registration Severino Cavalcante, in the city of Campina Grande on 25/11/2001 . Marriage and divorce, in general, and especially in Brazil were explored , and a whole literature of non-parametric and parametric methods with their probability distribution functions. Survival analysis is the term used to describe the statistical analysis When the study variable is the time from a well-defined initial time until occurs of the event of interest.

**Keywords:** Divorce; Kaplan-Meier; Parametrics Methods.



# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
<b>2</b>	<b>Fundamentação Teórica</b>	p. 13
2.1	Casamento . . . . .	p. 13
2.1.1	Casamento Civil . . . . .	p. 14
2.1.2	Casamento civil no Brasil . . . . .	p. 15
2.2	Divórcio . . . . .	p. 15
2.2.1	A evolução histórica do divórcio no Brasil . . . . .	p. 15
2.3	Taxa de casamentos no Brasil . . . . .	p. 16
2.4	Análise de Sobrevivência . . . . .	p. 17
2.4.1	Função de sobrevivência . . . . .	p. 17
2.4.2	Função de Risco . . . . .	p. 18
2.4.3	Falha ou morte . . . . .	p. 19
2.4.4	Censura . . . . .	p. 19
2.4.4.1	Tipos de Censura . . . . .	p. 19
2.5	Testes não paramétricos . . . . .	p. 20
2.6	O Estimador de Kaplan-Meier . . . . .	p. 21
2.7	Modelos Paramétricos . . . . .	p. 22
2.7.1	Distribuição Exponencial . . . . .	p. 22

2.7.2	Distribuição Weibull . . . . .	p. 24
2.7.2.1	Definição . . . . .	p. 25
2.7.3	Distribuição Gama . . . . .	p. 25
2.7.3.1	Definição . . . . .	p. 25
2.8	Método de Máxima Verossimilhança . . . . .	p. 26
<b>3</b>	<b>Aplicação</b>	p. 30
<b>4</b>	<b>Conclusão</b>	p. 36
	<b>Referências</b>	p. 37
	<b>Apêndice A – Apêndice</b>	p. 39
	<b>Anexo A – Anexo</b>	p. 40

# Lista de Figuras

1	Curva téorica do tempo de sobrevivência indicando o tempo mediano (Fonte: Carvalho et al. (2005)) . . . . .	p. 18
2	Representação gráfica da censura a direita. Fonte: (STRAPSSON, 2007)	p. 20
3	Distribuição exponencial com diferentes valores . . . . .	p. 24
4	Representação gráfica da distribuição gama Função acumulada (a) e função de distribuição (b). Fonte: (Portal Action, 2014) . . . . .	p. 26
5	Kaplan-Meier para representar o risco do divórcio . . . . .	p. 33
6	Kaplan-Meier para representar o tempo de sobrevivência do casamento . .	p. 34
7	Gráficos dos ajustes das distribuições paramétricas vs Kaplan-Meier . .	p. 35

# Lista de Tabelas

1	Dados de Matrimônios do 1 cartório de registro civil “Severino Galileu” da cidade de Campina Grande - PB, coletados por Albuquerque (2002). . .	p. 30
2	Estimativas de Kaplan-Meier para homens com idades até 24 anos (grupo 1) e acima de 24 anos (grupo 2). . . . .	p. 31
3	Teste logrank para o grupo de homens . . . . .	p. 31
4	Estimativas de Kaplan-Meier para mulheres com idades até 21 anos (grupo 1) e acima de 21 anos (grupo 2). . . . .	p. 32
5	Teste logrank para o grupo de mulheres . . . . .	p. 32
6	Teste da Razão de Verossimilhanças aplicado ao tempo até o divórcio .	p. 34
7	Valores estimados para a função de sobrevivência de Kaplan-Meier e das distribuições exponencial, lognormal e weibull. . . . .	p. 35

# 1 Introdução

Os primórdios do casamento se deve a uma sequência de eventos históricos, com a entrada do Estado nas relações, surgiu o casamento civil que é um contrato entre o estado e duas pessoas tradicionalmente com o objetivo de constituir família. No entanto, para resguardar a vontade de um ou dos dois entes do casamento o divórcio (do latim *divortium*, derivado de *divertere*, “separar-se”) é o rompimento legal e definitivo do vínculo de casamento civil. É uma das três maneiras de dissolver um casamento, além da morte de um dos cônjuges. O processo legal de divórcio pode envolver questões como atribuição de pensão de alimentos, regulação de poder paternal, relação ou partilha de bens, regulação de casa de morada de família, embora estes acordos sejam complementares ao processo principal.

A análise de sobrevivência é uma técnica em que a variável dependente é sempre o tempo até ocorrência de um determinado evento. Ela difere da análise estatística “clássica” devido a variável dependente ser o evento de interesse (tempo até a morte, cura, reincidente, etc...). Ela é uma das técnicas que mais cresceram nas últimas décadas.

Em estudos em que existem longos períodos de tempo (“follow-up”), os indivíduos sobre estudo podem não atingirem o tempo total previsto. A análise de sobrevivência considera as informações geradas por estes indivíduos, mesmo que eles não cheguem a desenvolver o evento de interesse. Isto pode ocorrer porque abandonaram o estudo, sofreram eventos não relacionados ao de estudo ou o estudo chegou ao seu término. Quando isto acontece, designa-se o indivíduo por censurado (Botelho, Silva e Cruz, 2009). Assim, censurado significa que o indivíduo não desenvolveu o evento até ao fim de observação no estudo (independentemente do motivo pelo qual tal lhe aconteceu ou do que lhe aconteceu posteriormente).

A grande vantagem da análise de sobrevivência é que permite utilizar a informação de todos os participantes até ao momento em que desenvolvem o evento ou são censurados enquanto a análise “clássica”, dado que estes doentes não estiveram todo o tempo em observação, têm que ser excluídos da análise, já que se desconhece quanto tempo demoraram a desenvolver o evento em causa. Assim a análise de sobrevivência é a técnica ideal para analisar respostas binárias (ter ou não ter um evento) em estudos longitudinais

que se caracterizam por tempo de seguimento diferente entre os indivíduos e perdas de follow-up.

O presente estudo trata da utilização de ferramentas estatísticas especialmente a análise de sobrevivência sobre números de divórcios na cidade de Campina Grande usando dados coletados no 1<sup>o</sup> cartório do registro civil Severino Cavalcante na cidade de Campina Grande em 25/11/2001 (no período de tempo estudado de 1995 à 2001). O objetivo principal deste trabalho consiste na busca de uma distribuição paramétrica que modele o tempo até o divórcio, como alternativa a função de sobrevivência utilizada para o estimado de kaplan-Meyer.

## 2 Fundamentação Teórica

Esta seção tem como objetivo retratar de forma clara e pontual os aspectos históricos referentes ao casamento e divórcio, bem como os aspectos teóricos a cerca dos modelos de sobrevivida por meio de livros, artigos, teses e dissertações.

### 2.1 Casamento

Os primórdios do casamento se deve a uma sequência de eventos históricos, em 392, o cristianismo foi proclamado religião oficial. Entre os anos de 965 e 1008 eram batizados e por tanto adeptos do cristianismo os reis da Dinamarca, Polônia, Rússia, Noruega e Suécia. A Europa pagã, mal batizada no ano 1000, apresentava uma concepção do casamento totalmente contrária à dos cristãos. Os vikings praticavam um casamento poligâmico, com uma esposa de primeiro escalão que tinha todos os direitos, e com esposas ou concubinas de segundo escalão, cujos filhos não tinham nenhum direito, a menos que a oficial fosse estéril, ou tivesse sido repudiada. As cerimônias de noivado organizavam a transmissão de bens, mas não havia casamento verdadeiro a não ser que tivesse havido união carnal. Na manhã da noite de núpcias, o esposo oferecia à mulher um conjunto muitas vezes bastante significativo de bens móveis. Ele era chamado de presente matinal (Morgengabe), que os juristas romanos batizaram de dote. Com a cristianização da Europa e elementos vindos do casamento pagão nasce em principio dos anos 1000 o casamento com uma face totalmente nova.

Durante o Sacro Império Romano Germânico - que sucedeu ao desaparecido Império Romano -, dirigido por Oto III de 998 a 1002, houve uma fabulosa transformação das sociedades urbanas romanas e das sociedades rurais germânicas e eslavas. As uniões entre homens e mulheres eram, então, o resultado complexo de renitências pagãs, de interesses políticos e de uma poderosa evangelização.

Na Idade Antiga o casamento era um acordo formal entre o noivo e o pai da noiva que implicava no pagamento de um dote por parte do pai e havia uma celebração religiosa domiciliar. Assim, esta forma de união conjugal não levava em conta a vontade da noiva nem o seu consentimento (COSTA, 2000).

Na Idade Média (séculos XI-XII) o casamento passa a ser um sacramento da Igreja constituindo um modelo conjugal cristão, no qual a indissolubilidade do casamento era exigida assim como a imagem de pureza da união. Não se dava importância ao amor no relacionamento e a validade do sacramento do matrimônio residia na fidelidade e em filhos em comum, portanto, o amor entre os cônjuges era considerado um resultado da união e não como base do relacionamento.

O casamento tinha por função ligar duas famílias e permitir que elas se perpetuassem, mais do que satisfazer o amor de duas pessoas. Um novo ideal de casamento constituiu-se aos poucos no Ocidente, em que se impõe aos cônjuges que se amem e que tenham expectativas a respeito do amor.

Desses dois fatos resultou o formato do casamento, em princípios do ano 1000, com uma face totalmente nova. Durante o Sacro Império Romano Germânico - que sucedeu ao desaparecido Império Romano -, dirigido por Oto III de 998 a 1002, houve uma fabulosa transformação das sociedades urbanas romanas e das sociedades rurais germânicas e eslavas. As uniões entre homens e mulheres eram, então, o resultado complexo de renitências pagãs, de interesses políticos e de uma poderosa evangelização.

### 2.1.1 Casamento Civil

O casamento civil é um contrato entre o estado e duas pessoas tradicionalmente com o objetivo de constituir família. A definição exata varia historicamente e entre as culturas mas até a pouco tempo na maioria dos países era uma união socialmente sancionada entre um homem e uma mulher (com ou sem filhos) mediante comunhão de vida e bens. Até ao século XIX o casamento era visto nas sociedades ocidentais (tal como acontece hoje em dia em muitos locais) meramente como um acordo comercial entre duas famílias sem que os dois intervenientes tivessem muito voto na matéria. O Romantismo veio alterar esta imagem e passou-se a existir o conceito de se casar por amor. Até o século XX era comum que o casamento fosse visto como algo indissolúvel (embora pudesse ser anulado) não havendo reconhecimento legal do divórcio. É crescente o número de países que reconhecem aos casais formados por dois homens ou duas mulheres, o acesso a este direito, inclusive no Brasil por decisão do Conselho Nacional de Justiça (CNJ), mais um indício da dinâmica do significado, assim como existem outros mecanismos legais de proteção da família de forma menos restritiva como a União de fato.



## 2.1.2 Casamento civil no Brasil

No dia 24 de janeiro de 1890 foi promulgado pelo Marechal Deodoro da Fonseca Chefe do Governo Provisório da então República dos Estados Unidos do Brasil, o Decreto nº 181, instituindo o casamento civil no país. No Brasil o casamento é regulamentado pelo Código civil. Ele é necessariamente monogâmico, e pode ser celebrado por casais heteroafetivos ou homoafetivos; via de regra, a idade mínima dos noivos (idade núbil) é de 16 anos. É um contrato bilateral e solene realizado entre as partes com o intuito de constituir família com uma completa comunhão de vida.

## 2.2 Divórcio

O divórcio (do latim *divortium*, derivado de *divertere*, “separar-se”) é o rompimento legal e definitivo do vínculo de casamento civil. É uma das três maneiras de dissolver um casamento, além da morte de um dos cônjuges. O processo legal de divórcio pode envolver questões como atribuição de pensão de alimentos, regulação de poder paternal, relação ou partilha de bens, regulação de casa de morada de família, embora estes acordos sejam complementares ao processo principal. Em algumas jurisdições não é exigida a invocação da culpa do outro cônjuge. Ainda assim, mesmo nos ordenamentos jurídicos que adaptaram o sistema do divórcio “sem culpa”, é tido em conta o comportamento das partes na partilha dos bens, regulação do poder paternal, e atribuição de alimentos.

### 2.2.1 A evolução histórica do divórcio no Brasil

Trata-se da evolução natural do direito brasileiro ao longo dos anos, conforme há mudanças sociais, necessita-se nova adequação da legislação acerca do tema para que os anseios da sociedade sejam atendidos. O tema casamento x divórcio sofre sérias mudanças ao longo das décadas, deixando de lado a legislação que resguardava o primeiro, flexibilizando assim, as leis para obtenção do segundo.

Ao longo do século XX a família passou por grandes transformações na seara jurídica, deixando de ser a família patriarcal, instituída em torno do patriarca, responsável por todos os integrantes do núcleo familiar, para ser concebida como a família plural, fundada no amor e no afeto.

Este novo cenário possibilitou que o matrimônio, antes indissolúvel, passasse por grande evolução, o que contribuiu para o fortalecimento da mulher, enquanto sujeito de

direito, e para o fortalecimento da própria instituição familiar.

No Brasil o advento da Constituição da Republica de 1988 e considerado um verdadeiro divisor de águas, pois deu novos contornos a entidade familiar, além de ter contribuído para o processo de despatrimonialização do Direito de Família, para a sua repersonalização, ou seja, fundar seus objetivos e institutos da pessoa, motivo pelo qual os princípios constitucionais, em especial o princípio da dignidade da pessoa humana, ganharam importância.

O advento do Código Civil de 2002, que disciplinou o Divórcio em sua totalidade, mantendo o procedimento prévio, denominada separação judicial, ao lado do divórcio, que podia se dar por conversão ou de forma direta, observados os prazos legais. A Emenda Constitucional nº 66/2010 que transformou significativamente o instituto, suprimindo o instituto da separação judicial, afastando as discussões do elemento culpa e excluindo a exigência de lapso temporal.

## 2.3 Taxa de casamentos no Brasil

Segundo o Instituto Brasileiro de Geografia e Estatística (2012) a taxa de nupcialidade no Brasil foi de 6,9% no ano de 2012, sendo registrados neste ano 1041440 casamentos, 1,4% a mais que no ano anterior. Deste total, 1.040.473 foram de cônjuges de 15 anos ou mais, valor que manteve a taxa de nupcialidade legal estável em relação a 2011, com 6,9 casamentos para mil habitantes de 15 anos ou mais de idade. Entre 2002 e 2012, observa-se tendência de elevação da taxa de nupcialidade legal, embora os patamares alcançados sejam bem inferiores aos observados na década de 1970, quando era de 13%.

Em 2012, 2,9% dos cônjuges brasileiros do sexo masculino eram menores de 20 anos de idade, ao passo que entre as mulheres 12,4% dos cônjuges estavam nessa faixa etária. A maior taxa de nupcialidade para as mulheres está no grupo de 20 a 24 anos (30,0%), valor próximo ao de 2007 (30,2%), mas com tendência de declínio. De 15 a 19 anos, a taxa em 2012 (15,3%) foi inferior à de 2002 (15,7%). Já entre 25 e 29 anos, houve contínua elevação no período (de 21,2% em 2002 para 29,0% em 2012), indicando o aumento da idade média das mulheres ao casar. As taxas de nupcialidade das mulheres são maiores que as dos homens apenas nos dois grupos etários mais jovens. Entre as mulheres de 15 a 19 anos, a taxa foi de 15,3%, ao passo que, entre os homens, ficou em 3,5%. A partir dos 60 anos, as taxas para homens (3,8%) são mais que o dobro que das taxas femininas (1,1%).

## 2.4 Análise de Sobrevivência

A análise de sobrevivência consiste em uma coleção de procedimentos estatísticos para análise de dados relacionados ao tempo decorrido desde um tempo inicial, pré-estabelecido, até a ocorrência de um evento de interesse. As principais características das técnicas de análise de sobrevivência são sua capacidade de extrair informações de dados censurados, ou seja, observações para as quais, no final de um determinado período de tempo pré-estabelecido de acompanhamento, o problema de pesquisa não foi observado, além de levar em consideração o tempo para ocorrência dos eventos. De maneira geral, um tempo censurado corresponde ao tempo decorrido entre o início e o término do estudo do evento de interesse para ele. (LOUZADA et al., 2012)

Na análise de sobrevivência, o comportamento da variável aleatória tempo de sobrevivência,  $T \geq 0$ , pode ser expresso por meio de várias funções matematicamente equivalentes, tais que, se uma delas é especificada, as outras podem ser derivadas. Essas funções são: a função de densidade de probabilidade,  $f(t)$ , a função de sobrevivência,  $S(t)$ , e a função de risco,  $h(t)$ . Essas três funções são utilizadas na prática para descrever diferentes aspectos apresentados pelo conjunto de dados.

### 2.4.1 Função de sobrevivência

A função de sobrevivência é uma das principais funções probabilísticas usadas para descrever dados de tempo de sobrevivência. Tal função é definida como a probabilidade de não ser observado o evento de interesse para um indivíduo sobreviver ao tempo  $t$ , ou seja, a probabilidade de um indivíduo sobreviver ao tempo  $t$  sem o evento. Em termos probabilísticos esta função é dada por:

$$S(t_1) = \frac{n[1 - h(t_1)]}{n} = 1 - h(t_1)$$

Tempos de sobrevivência são também referenciados como dados de sobrevivência, observações exatas ou ainda observações não censuradas.

Outra forma é dada por Carvalho et al. (2005)

$$S(t) = P(T > t)$$

Probabilidade de um indivíduo sobreviver por mais do que um determinado tempo  $t$ , ou

por no mínimo um tempo igual a  $t$ .

A probabilidade de sobreviver até o tempo  $t$  (um evento ocorrer até o tempo  $t$ )

$$F(t) = P(T \leq t)$$

Assim,

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

Na figura 1, ilustra-se a função de sobrevivência (eixo das abscissas) versus o tempo de sobrevivência (eixo das ordenadas). Nesta figura quando o tempo é igual a 0 a função de sobrevivência apresenta valor igual a 1, quando o limite  $S(t)$  decresce, a função assume  $S(\infty) = 0$  quando  $t \rightarrow \infty$

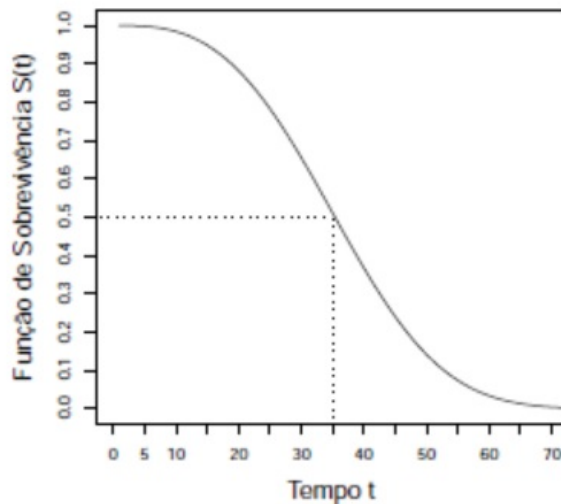


Figura 1: Curva teórica do tempo de sobrevivência indicando o tempo mediano (Fonte: Carvalho et al. (2005))

## 2.4.2 Função de Risco

A função de risco  $\lambda(t)$  ou  $h(t)$  é definida segundo Carvalho et al. (2005) e Colosimo e Giolo (2006) o risco instantâneo de um indivíduo sofrer o evento no intervalo de tempo  $(t, t + \varepsilon)$ , dado que ele sobreviveu até o tempo  $t$ . Carvalho et al. (2005) afirmam que por se tratar de incremento de tempo, esta função expressa o risco instantâneo de ocorrência de um evento dado que o evento não tenha ocorrido,

$$\lambda(t) = \lim_{\varepsilon \rightarrow 0} \frac{P[(t \leq T \leq t + \varepsilon | T \geq t)]}{\varepsilon}$$

essa função recebe vários nomes na literatura sendo um dos mais comuns função taxa de falha. A função de risco está relacionada com a função de sobrevivência por meio das expressões:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

e

$$\lambda(t) = \frac{-d\ln(S(t))}{dt}$$

quando o risco aumenta, a probabilidade de sobreviver diminuí.

A função risco acumulado  $\Lambda(t)$  mede o risco de ocorrência do evento até um determinado tempo  $t$ . Como  $T$  é uma variável aleatória contínua, esta função é a soma infinitesimal e será calculada por

$$\Lambda(t) = \int_0^t \lambda(u)du$$

ou

$$\Lambda(t) = -\ln S(t)$$

### 2.4.3 Falha ou morte

É a ocorrência de um determinado evento, que pode ou não ser pré-estabelecido no início da pesquisa. Por exemplo, uma falha pode ser a morte de um ser em estudo ou uma recaída, mas também pode ser considerada como a melhora do quadro clínico do paciente.

### 2.4.4 Censura

Segundo Colosimo;Giolo, 2006 censura é o registro parcial do tempo de falha, devido à perda ou retirada de um elemento do estudo. Por exemplo, em um estudo para se verificar o tempo até a reicidiva da doença, a censura será considerada se os indivíduos em estudo não apresentarem a reicidiva da doença até o fim do período de estudo, ou se saírem do estudo por causas outras que não a de interesse, morte ou abandono do tratamento.

#### 2.4.4.1 Tipos de Censura

Existem três tipos de censuras utilizadas em estudos de sobrevivência:

- i) A censura Tipo I é aquela, onde o teste será terminado após um período pré-estabelecido de tempo.

Neste caso o estudo é iniciado com um número fixo de pessoas (ou quaisquer seres vivos). O tempo de sobrevivência (desde o início do experimento até a falha) é registrado, ou seja, é obtida uma observação exata ou não censurada. O tempo de sobrevivência dos seres que ainda não morreram até o término do estudo não será conhecido, por isso é registrado o período mínimo observado, isto é, a observação censurada, existindo a censura a direita que aquela que os indivíduos são censurados de acordo com o tempo e a censura a esquerda em que antes do estudo ter inicio a censura já ocorreu. Na Figura 2 tem uma representação da censura a direita.

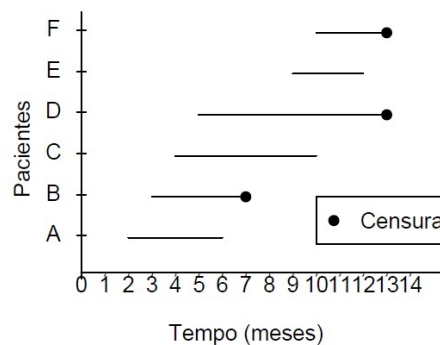


Figura 2: Representação gráfica da censura a direita. Fonte: (STRAPSSON, 2007)

- ii) A Censura Tipo II é aquela, onde o teste será terminado após ter ocorrido a falha em um número pré-estabelecido de seres sob teste.

Trata-se de esperar até que uma parte pré-estabelecida dos seres morra. Dos sobreviventes são então extraídas as observações censuradas.

- iii) Na Censura Tipo III o período de estudo é fixado e os seres entram no estudo em diferentes tempos durante aquele período. Alguns podem morrer antes do fim do estudo e os tempos de sobrevivência exatos deles são conhecidos. Outros podem ser retirados antes do final do estudo e assim, são perdidas as respostas. E ainda tem aqueles que podem estar vivos no final do estudo.

## 2.5 Testes não paramétricos

Segundo Ferreira (2010) os testes de hipótese recebem a designação de testes paramétricos se satisfazem simultaneamente as seguintes condições:

a) Os testes incidem explicitamente sobre um parâmetro de uma ou mais populações (por exemplo, sobre a média ou valor esperado, ou sobre a variância);

b) A distribuição de probabilidades da estatística de teste pressupõe uma forma particular das distribuições populacionais de onde as amostras foram recolhidas. Por exemplo, a distribuição da estatística de teste do teste t-Student para comparar as médias de duas amostras pressupõe que as amostras foram retiradas de uma população que se distribui segundo uma função de probabilidades normal, e além disso pressupõe também que as variâncias das duas amostras são homogêneas (a fórmula de Welsh para calcular os graus de liberdade do teste de comparação de duas amostras cujas variâncias não são homogêneas resulta se a violação deste pressuposto não é muito acentuada. Se as variâncias são muito diferentes, é preferível usar um teste não paramétrico.

c) Os erros ou resíduos  $\varepsilon_i$  (tal que  $x_i = \mu + \varepsilon_i$ ) têm distribuição normal;

d) Os erros ou resíduos  $\varepsilon_i$  têm variância finita e constante  $\sigma^2$  ;

e) Os erros ou resíduos  $\varepsilon_i$  são independentes.

Assim, se algum destes pressupostos é violado, então os testes tradicionais vistos anteriormente não têm rigor estatístico, e deverão ser evitados, e em sua substituição dever-se-ão utilizar testes que não exigem o cumprimento de tais pressupostos. Estes testes designam-se por testes não paramétricos. Os testes não paramétricos não estão condicionados por qualquer distribuição de probabilidades dos dados em análise, sendo também designados por *distribution-free tests*. Tal como não é estatisticamente rigorosa a utilização de testes paramétricos quando não se cumprem os pressupostos necessários, também deverá ser evitada a utilização dos testes não paramétricos em situações em que prevalecem as condições de utilização dos testes paramétricos, pois estes (paramétricos) são mais poderosos que os testes não paramétricos (FERREIRA, 2010).

Para os testes paramétricos ou não paramétricos, qualquer teste de hipóteses só tem validade estatística se as amostras sobre as que estão a ser aplicados forem aleatórias.

## 2.6 O Estimador de Kaplan-Meier

Na prática, os conjuntos de dados amostrais de tempos de falha apresenta censuras, o que requer técnicas estatísticas especializadas para acomodar a informação contida nestas observações. A observação censurada informa que o tempo até a falha é maior do que aquele que foi registrado. Nesta seção é apresentado o conhecido estimador de Kaplan-

Meier para a função de sobrevivência, que é, sem dúvida, o mais utilizado em estudos clínicos e vem ganhando cada vez mais espaço em estudos de confiabilidade.

O estimador não-paramétrico de Kaplan-Meier, proposto por Kaplan e Meier (1958) para estimar a função de sobrevivência, é também chamado de estimador limite-produto. Ele é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\widehat{S}_t = \frac{\text{número de observações que não falharam até o tempo } t}{\text{número total de observações no estudo}}$$

$\widehat{S}_t$  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 um certo tempo  $t$ , o tamanho do degrau fica multiplicado pelo número de empates. O estimador de Kaplan-Meier pode ser apresentado após algumas definições.

- i)  $t_1 < t_2 < \dots < t_k$ , os  $k$  tempos distintos e ordenados de falha,
- ii)  $d_j$  o número de falhas em  $t_j, j = 1, \dots, k$  e
- iii)  $n_j$  o número de indivíduos sob risco em  $t_j$ , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a  $t_j$ .

Deste modo, o estimador é:

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

## 2.7 Modelos Paramétricos

Nesta seção serão abordados os modelos probabilísticos utilizados para o desenvolvimento deste trabalho, as expressões para suas funções empíricas, sobrevivência e risco. Tais modelos são importantes para descrever de forma adequada certas variáveis quando estas se tratam do “tempo até a falha”.

### 2.7.1 Distribuição Exponencial

Esta é uma distribuição que se caracteriza por ter uma função de taxa de falha constante. A distribuição exponencial é a única com esta propriedade. Ela é considerada



uma das mais simples em termos matemáticos. Esta distribuição tem sido usada extensivamente como um modelo para o tempo de vida de certos produtos e materiais. Ela descreve adequadamente o tempo de vida de óleos isolantes e dielétricos, entre outros.

A variável aleatória  $T$  tem distribuição exponencial com parâmetro  $\lambda$ ,  $\lambda > 0$ , se tiver função densidade de probabilidade dada por:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{se } t \geq 0 \\ 0 & \text{se } t < 0 \end{cases}$$

em que, o parâmetro  $\lambda > 0$  é o tempo médio de vida e  $t$  é um tempo de falha. O parâmetro deve ter a mesma unidade do tempo da falha  $t$ . Isto é, se  $t$  é medido em horas,  $\lambda$  também será medido em horas. A função de distribuição acumulada  $F(t)$  é dada por

$$F(t) = \int_0^t f(s)ds = \begin{cases} 1 - e^{-\lambda t} & \text{se } t \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Com a função de probabilidade  $f(t)$  e a função de probabilidade acumulada  $F(t)$  podemos calcular o valor da função de sobrevivência  $S(t)$  que é definida como a probabilidade de um indivíduo sobreviver até um certo tempo  $t$ , sem o evento. Sendo uma das principais funções probabilísticas usadas para descrever dados de tempo de sobrevivência, definida por:

$$S(t) = P(T > t) = 1 - F(t) = \int_0^t f(s)ds, \quad t \geq 0 \quad (2.1)$$

Agora substituindo  $F(t)$  calculados acima temos que

$$1 - F(t) = 1 - (1 - e^{-\lambda t}) = 1 - 1 + e^{-\lambda t} = e^{-\lambda t} = S(t)$$

A função de risco, ou taxa de falha descreve a forma com que a taxa de falha muda com o tempo, ou seja, demonstra o risco do indivíduo falhar no tempo. A função de risco pode ser definida, em termos da função de distribuição  $F(t)$  e da função de densidade de probabilidade  $f(t)$ , porém podemos substituir  $1 - F(t)$  por  $S(t)$  da seguinte forma:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0 \quad (2.2)$$

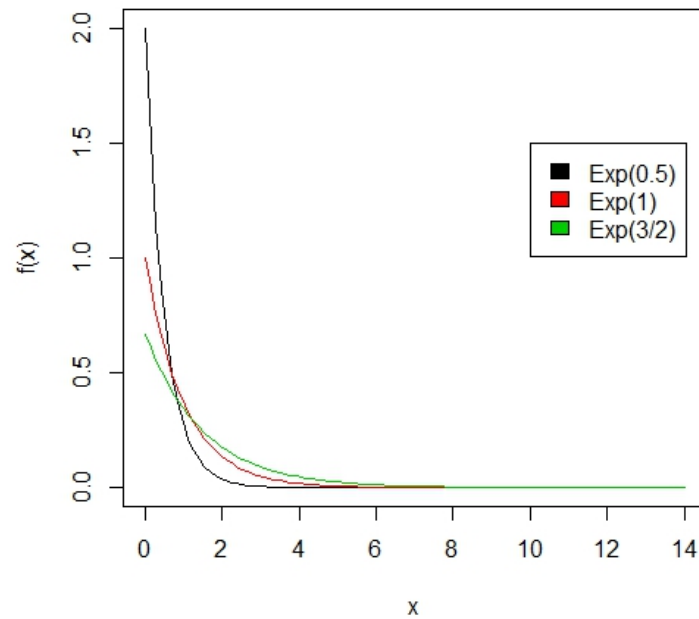


Figura 3: Distribuição exponencial com diferentes valores

substituindo  $f(t)$  e  $S(t)$  pelos resultados obtidos logo acima, obtemos

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \cdot e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Observação: Notem que a função exponencial, na verdade, é um caso particular da função Gama, pois se  $X \sim \text{Exp}(\lambda)$ , então  $X \sim \text{Gama}(1, \lambda)$

Na figura 3 tem-se a distribuição exponencial com parâmetros  $\lambda = 1/2, 1$  e  $3/2$ .

## 2.7.2 Distribuição Weibull

A distribuição Weibull foi proposta originalmente por W. Weibull (1954) em estudos relacionados ao tempo de falha devido a fadiga de metais. Ela é frequentemente usada para descrever o tempo de vida de produtos industriais. A sua popularidade em aplicações práticas deve-se ao fato dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de taxa de falha é monótona. Isto é, ou ela é crescente ou decrescente ou constante. Ela descreve adequadamente a vida de mananciais, componentes eletrônicos, cerâmicas, capacitores e dielétricos.

### 2.7.2.1 Definição

Uma variável aleatória  $X$  tem distribuição Weibull se tiver função densidade de probabilidade dada por:

$$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp \left[ - \left( \frac{x}{\beta} \right)^\alpha \right], & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Sua função de distribuição acumulada é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - \exp \left[ - \left( \frac{x}{\beta} \right)^\alpha \right] & \end{cases}$$

As funções de risco e de sobrevivência são, respectivamente,

$$S(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\} \quad (2.3)$$

e

$$\lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \quad (2.4)$$

para  $t \geq 0$ ,  $\alpha$  e  $\gamma > 0$ .

## 2.7.3 Distribuição Gama

A distribuição gama é uma das mais gerais distribuições, pois diversas distribuições são caso particular dela como por exemplo a exponencial, a qui-quadrado, entre outras. Essa distribuição tem como suas principais aplicações à análise de tempo de vida de produtos.

### 2.7.3.1 Definição

Uma variável aleatória  $X$  tem distribuição Gama com parâmetros  $\alpha$  (também denominado parâmetro de forma) e  $\beta$  (parâmetro de taxa), denotando-se  $X \sim \text{Gama}(\alpha, \beta)$ , se sua função densidade é dada por

$$f(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

Os gráficos aproximados das funções de distribuição e de densidade para  $\alpha = 2$  e  $\beta = 0,5$  são apresentados, respectivamente, na seguinte figura

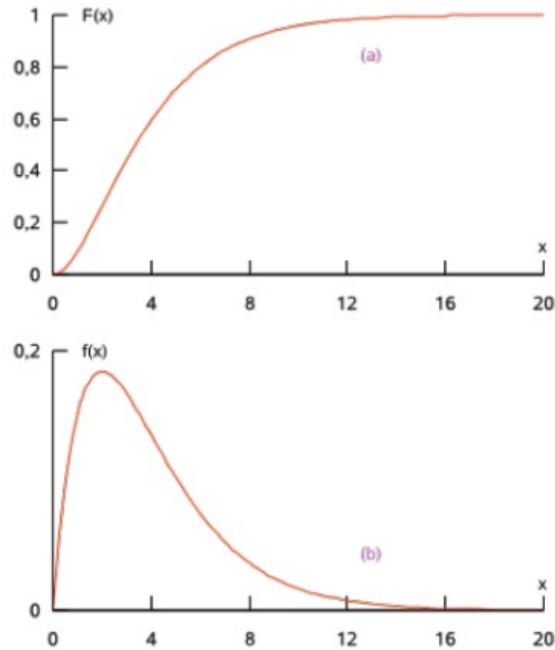


Figura 4: Representação gráfica da distribuição gama Função acumulada (a) e função de distribuição (b). Fonte: (Portal Action, 2014)

## 2.8 Método de Máxima Verossimilhança

O método de máxima verossimilhança trata o problema de estimação da seguinte forma: baseado nos resultados obtidos pela amostra, devemos determinar qual a distribuição, dentre todas aquelas definidas pelos possíveis valores de seus parâmetros, com maior possibilidade de ter gerado tal amostra. Em outras palavras, se por exemplo a distribuição do tempo de falha é a de Weibull, para cada combinação diferente de  $\alpha$  e  $\delta$  tem-se diferentes distribuições de Weibull. O estimador de máxima verossimilhança escolhe aquele par de  $\alpha$  e  $\delta$  que melhor explica a amostra observada. A seguir discute-se as idéias do método de máxima verossimilhança para conceitos matemáticos a partir dos quais será possível obter estimadores para os parâmetros. Suponha uma amostra de observações  $t_1, \dots, t_n$  de uma certa população de interesse. Considere inicialmente que todas as observações são não-censuradas. A população é caracterizada pela sua função de densidade  $f(t)$ . Por exemplo, se  $f(t) = \alpha \exp(-\alpha t)$ , significa que as observações vem de uma distribuição Exponencial com um parâmetro a ser estimado. A função de verossimilhança para um parâmetro genérico é dada por

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta) \quad (1)$$

Note na expressão acima que  $\theta$  pode representar um único parâmetro ou um conjunto de parâmetros. Por exemplo, no modelo Log-normal temos  $\theta = (\mu, s)$ . A tradução em termos matemáticos para a frase “a distribuição que melhor explica a amostra observada” é achar o valor  $\theta$  que maximiza a função  $L(\theta)$ . Isto é, achar o valor de  $\theta$  que maximiza a probabilidade da amostra observada ocorrer. A função de verossimilhança  $L(\theta)$  mostra que a contribuição de cada observação não-censurada é a sua função de densidade. Por outro lado, a contribuição de cada observação censurada não é a sua função de densidade. Essas observações somente nos informam que o tempo de falha é maior do que o tempo de censura observado e portanto, sua contribuição para  $L(\theta)$  é a sua função de confiabilidade  $R(t)$ . As observações podem então ser divididas em dois conjuntos, as  $r$  primeiras são as não-censuradas (1, 2, ...,  $r$ ), e as  $n - r$  seguintes são as censuradas ( $r + 1, r + 2, \dots, n$ ). Com isso, a função de verossimilhança assume a seguinte forma

$$L(\theta) = \prod_{i=1}^r f(t_i; \theta) \prod_{i=r+1}^n R(t_i; \theta) \quad (4.2.1.2)$$

A expressão (1) para a verossimilhança é válida para os mecanismos de censura do tipo I e II sob a suposição de que o mecanismo de censura é não-informativo, ou seja, não carrega informações sobre os parâmetros, a mesma vale também para o mecanismo do tipo aleatório. Essa suposição é razoável em estudos de durabilidade e é sempre conveniente trabalhar com o logaritmo da função de verossimilhança. Os estimadores de máxima verossimilhança são os valores  $\theta$  que maximizam  $L(\theta)$  ou equivalente  $\log(L(\theta))$ . Eles são encontrados resolvendo o sistema de equações

$$U(\theta) = \frac{\partial \log(L(\theta))}{\partial \theta} = 0$$

Outra abordagem é dada por Carvalho et al. (2005), em que a função de verossimilhança é dada por

$$L \propto \prod_{i \in O} f(t_i) \prod_{i \in D} S(t_i)$$

em que  $O$  é o conjunto de observações que sofreram evento e  $D$  é o conjunto de observações que foram censuradas a direita. Carvalho et al. (2005) afirmam que quando existe censura

a esquerda sabe-se que apenas o tempo de sobrevivência exato é menor que o observado e conseqüentemente a contribuição dessas observações é dada pela função de distribuição acumulada

$$F(t) = 1 - S(t)$$

Assim,

$$L \propto \prod_{i \in O} f(t_i) \prod_{i \in D} S(t_i) \prod_{i \in \varepsilon} (1 - S(t_i))$$

em que  $\varepsilon$  é o conjunto de observações que foram censuradas a esquerda. Segundo Carvalho et al. (2005) no caso de censuras intervalares é necessário incluir um proditório a mais na formula da verossimilhança, de modo que a probabilidade de ocorrência do evento dentro de um intervalo de tempo  $(t^-, t^+)$  assim,

$$L \propto \prod_{i \in O} f(t_i) \prod_{i \in D} S(t_i) \prod_{i \in \varepsilon} (1 - S(t_i)) \prod_{i \in I} [S(t_i^-) - S(t_i^+)]$$

sendo  $I$  o conjunto de observações que sofreram censuras intervalares.

Para exemplificar a técnica de obtenção de estimadores por máxima verossimilhança demonstra-se o uso com a distribuição exponencial para uma variável aleatória  $T$  supondo  $T$  positiva e determinando o tempo até ocorrência de um evento de interesse e com a suposição adicional de censura apenas a direita, uma adaptação da função de verossimilhança é dada por:

$$L = \prod_{i=1}^n (f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}),$$

assim,

$$\begin{aligned} L &= \prod_{i=1}^n (\alpha \exp(-\alpha t_i)^{\delta_i} \exp(-\alpha t_i)^{1-\delta_i}) \\ l &= \ln L = \sum_{i=1}^n \delta_i (\ln \alpha - \alpha t_i) + (-\delta_i) (-\alpha t_i) \\ &= \sum_{i=1}^n \delta \ln \alpha - \alpha t_i \end{aligned}$$

O estimador de Máxima Verossimilhança para  $\alpha$  é dado pela derivada parcial de  $\frac{\partial l}{\partial \alpha} = 0$ , assim,

$$\hat{\alpha} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i} = \frac{\text{total de Eventos}}{\sum_{i=1}^n t_i}$$

Carvalho et al. (2005) afirmam que o parâmetro  $\alpha$  é calculado como número total de eventos dividido pela soma de todos os tempos observados, o que equivale a calcular o número de eventos por pessoa tempo, assim o risco é considerado constante, quando considerarmos o tempo de sobrevivência distribuído como uma distribuição exponencial que possui a propriedade da falta de memória.

### 3 Aplicação

Para a realização deste trabalho utilizou-se os dados coletados por Albuquerque (2002). Os dados foram coletados no 1 cartório de registro civil “Severino Cavalcante” na cidade de Campina Grande - PB em 25 de novembro de 2001. Tratam-se de dados coletados por meio de uma amostra de 44 casamentos que tiveram como data de início a partir de 01 de Julho de 1995, para os quais foi adotada a censura tipo I a direita, na qual a coleta das informações baseou-se em uma amostra aleatória simples sem reposição, para uma população com suposição de ser finita, composta de 600 casais ( $N = 600$ ) casados em 1995. O período de acompanhamento dos casais foi de 6,33 anos (seis anos e quatro meses), considerando censurado o casal que não tiver o evento de interesse (divorcio) no período estudado. Albuquerque (2002), detalha o procedimento de amostragem com um erro de 5,6% e coeficiente de confiança de 96% originando o tamanho da amostra calculado. Na amostra calculada houveram dois grupos separados pela mediana da idade calculadas em 21 anos para mulheres e 24 anos para homens, EXPLICAÇÃO DOS GRUPOS E DOS FATORES. Todas as análises foram feitas utilizando-se o pacote *survival* do R (R Development Core Team, 2014).

Na Tabela 1 tem-se uma amostra dos dados originais obtidos de Albuquerque (2002):

Tabela 1: Dados de Matrimônios do 1 cartório de registro civil “Severino Galileu” da cidade de Campina Grande - PB, coletados por Albuquerque (2002).

	tempo	evento	ID_M	ID_F	fator1	fator2
1	6,33	0	22	23	grupo1	grupo2
2	6,33	0	52	22	grupo2	grupo2
3	6,33	0	23	22	grupo1	grupo2
4	6,33	0	24	21	grupo1	grupo1
5	6,33	0	23	21	grupo1	grupo1
6	6,33	0	25	29	grupo2	grupo2

ID\_M - Idade Masculina; ID\_F - Idade Feminina

Na Tabela 2 tem-se as estimativas de sobrevivência de cada grupo obtidas por meio da função de sobrevivência de Kaplan-Meier para o fator 1 (homens). Percebe-se a ocorrência da primeira falha foi no tempo 4,25 (quatro anos e três meses) para homens com até 24 anos, os erros-padrão foram pequenos o que indica a um bom ajuste do estimador de



Kaplan-Meier, os intervalos de confiança para a sobrevivência estimada foram de baixa amplitude indicando boa precisão nas estimativas deste estimador. Percebe-se também que após 5,16 (5 anos e dois meses) a probabilidade de um casamento sobreviver fica abaixo de 80%. Para homens com idade acima de 24 anos percebe-se que o primeiro divórcio aconteceu mais cedo por volta de 3 anos e nove meses (3,83), porém após 5 anos e dois meses (5,16).

Foi aplicado o teste logrank (Tabela 3) para verificar se os grupos de idade tinham o mesmo comportamento no que se refere a sobrevivência estimada, ou se apresentavam comportamento diferente quanto ao risco de divórcio. A estatística de teste  $\chi^2 = 0,10$  com um grau de liberdade e um Valor P de 0,707, indicando alta probabilidade de se cometer um erro ao rejeitar  $H_0$ , ou seja, não existe diferença entre os grupos de idade pesquisados para o fator 1 (homens).

Tabela 2: Estimativas de Kaplan-Meier para homens com idades até 24 anos (grupo 1) e acima de 24 anos (grupo 2).

fator1=grupo1						
tempo	n.risco	n.evento	sobrevivência	Erro Padrão	Lim. inf. 95% IC	Lim. sup. 95% CI
4.25	24	1	0,958	0,0408	0,882	1,000
4.42	23	1	0,917	0,0564	0,813	1,000
4.50	22	2	0,833	0,0761	0,697	0,997
5.16	20	2	0,750	0,0884	0,595	0,945
fator1=grupo2						
3.83	20	1	0,95	0,0487	0,859	1,000
4.25	19	1	0,90	0,0671	0,778	1,000
5.16	18	1	0,85	0,0798	0,707	1,000
5.92	17	1	0,80	0,0894	0,643	0,996

Tabela 3: Teste logrank para o grupo de homens

	N	Observado	Esperado	$(O - E)^2/E$	$(O - E)^2/V$
fator1=grupo1	24	6	5,42	0,0630	0,141
fator1=grupo2	20	4	4,58	0,0744	0,141

Na Tabela 4 tem-se os resultados obtidos pelo estimador de Kaplan-Meier para o fator 2 (mulheres) para o grupo 1 (mulheres com idade até 21 anos) e grupo 2 (mulheres com idade maior que 21 anos).

Tabela 4: Estimativas de Kaplan-Meier para mulheres com idades até 21 anos (grupo 1) e acima de 21 anos (grupo 2).

fator2=grupo1						
tempo	n.risco	n.evento	Sobrevivência	Erro padrão	Lim. inf. 95% IC	Lim. sup. 95% IC
4,25	23	1	0,957	0,0425	0,877	1,000
4,50	22	1	0,913	0,0588	0,805	1,000
5,16	21	2	0,826	0,0790	0,685	0,996
5,92	19	1	0,783	0,0860	0,631	0,971
fator2=grupo2						
3,83	21	1	0,952	0,0465	0,866	1,000
4,25	20	1	0,905	0,0641	0,788	1,000
4,42	19	1	0,857	0,0764	0,720	1,000
4,50	18	1	0,810	0,0857	0,658	0,996
5,16	17	1	0,762	0,0929	0,600	0,968

Conforme o que ocorreu com homens percebe-se que para mulheres com idade até 21 anos (Tabela 4) o divórcio aconteceu primeiro que para as mulheres maiores de 21 anos, o que pode indicar a principio alguma relação entre idade e tempo de sobrevivência do matrimônio. Após 5,16 anos (cinco anos e dois meses) os casamentos tem uma probabilidade um pouco maior que 80% de sobreviver ao divórcio. Para o grupo 2 (mulheres maiores de 21 anos) essa probabilidade é menor de 80% para o tempo de 5,16 (cinco anos e dois meses). O teste logrank aplicado para mulheres afim de verificar se existia diferença entre os comportamentos das curvas de sobrevivência estimadas por Kaplan-Meier obteve um  $\chi^2 = 0,10$  com 1 grau de liberdade e um valor p de 0,782, o que indica que não existe diferenças entre as curvas de sobrevivência estimadas (Tabela 5).

Tabela 5: Teste logrank para o grupo de mulheres

	N	Observado	Experado	$(O - E)^2/E$	$(O - E)^2/V$
fator2=grupo1	23	5	5,43	0,0340	0,0765
fator2=grupo2	21	5	4,57	0,0404	0,0765

Pela Figura 5 percebe-se que o risco do divórcio aumenta consideravelmente após 4 anos de casados (para homens fator 1), devido não se ter coletado outras variáveis não é possível afirmar o que pode aumentar o risco de divórcio, tendo em vista que a idade dos cônjuges não alterou significamente as curvas de sobrevivência e risco. Porém sabe-se que alguns fatores tais como jornada de trabalho extensa, profissão, condições socioeconómico podem vir a alterar o tempo de sobrevivência dos casamentos, devido o comportamento ser similar para o fator 2 (mulheres) optou-se por não apresentar o gráfico de Kaplan-Meier.

As funções de sobrevivência estimadas pelo método de Kaplan-Meier indicam que após 3,5 anos a probabilidade de sobrevida do casamento diminuí(fator 1 homens), percebe-se

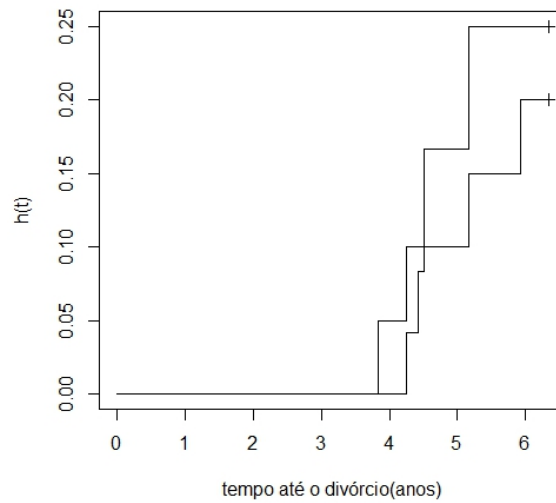


Figura 5: Kaplan-Meier para representar o risco do divórcio

também que as funções estimadas para este estudo não tendem a chegar a zero (FIGURA 6) com probabilidade 1 de divórcio, isto sugere que métodos de sobrevivência que consideram situações em que a função estimada nunca chega a zero podem ser considerados, conforme Taconeli (2013) quando uma população apresenta, após um período representativo de tempo, uma quantidade expressiva de observações censuradas, pode-se suspeitar que exista uma fração de indivíduos que não é susceptível ao evento de interesse. Diz-se então que esses indivíduos são "imunes", e que o conjunto de dados ao qual eles pertencem possui uma fração de cura. Os modelos de cura assumem implicitamente que todos os indivíduos que apresentaram o evento de interesse pertencem a uma população homogênea, optou-se por não apresentar a figura para o grupo feminino por esta ser similar a 6.

Após aplicar o estimador de sobrevivência de Kaplan-Meier desejou-se saber qual distribuição paramétrica (lognormal, exponencial, weibull) modelava o tempo até o divórcio. O intuito é uma vez sabendo qual distribuição paramétrica modela os dados em estudo é possível com a coleta de covariáveis relacionar os fatores que implicam no tempo de sobrevivência. Colosimo e Giolo (2006) afirmam que o conhecimento da correta distribuição de probabilidade para modelar o tempo de falha permite uma análise estatística mais precisa.

Na Tabela 6 foi aplicado o teste da razão de verossimilhanças para verificar qual modelo probabilístico se adequava aos dados, constatou-se pelo valor P (todos  $< 0,001$ ) que nenhuma das distribuições sugeridas se adequou bem aos dados. A ideia do teste

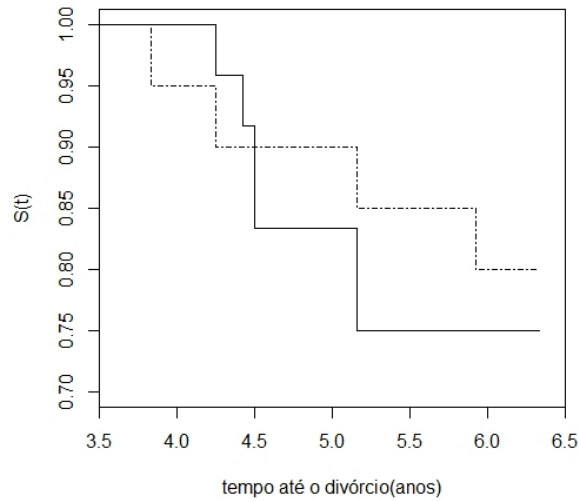


Figura 6: Kaplan-Meier para representar o tempo de sobrevivência do casamento

de razão de verossimilhanças é procurar uma distribuição parcimoniosa que descreva os dados tão bem quanto a distribuição gama generalizada.

Tabela 6: Teste da Razão de Verossimilhanças aplicado ao tempo até o divórcio

Modelo	$Log(L(\theta))$	TRV	Valor P
Gama Generalizado	-16,80	-	-
Exponencial	-42,67	$2(42,67-16,80)=51,47$	0,001
Weibull	-36,77	$2(36,77-16,80)=39,94$	0,001
Log Normal	-35,82	$2(35,82-16,80)=38,04$	0,001

Percebe-se na Tabela 7 que nenhuma das distribuições consideradas teve valores estimados para a função de sobrevivência próximos dos valores obtidos pelo estimador de Kaplan-Meier, corroborando o Teste da Razão de Verossimilhanças. Isto indica que das distribuições consideradas não se pode tirar conclusões acerca do tempo de falha, porém outras distribuições não consideradas e/ou métodos como o de regressão de cox podem vir a ter resultados melhores que estas distribuições paramétricas. Na figura 7 tem-se os gráficos de ajuste das distribuições com as estimativas de Kaplan - Meier.

Tabela 7: Valores estimados para a função de sobrevivência de Kaplan-Meier e das distribuições exponencial, lognormal e weibull.

	tempo	$\hat{S}_t$	$\hat{S}_{te}$	$\hat{S}_{tw}$	$\hat{S}_{tl}$
1	3,83	0,98	0,86	0,22	0,82
2	4,25	0,93	0,85	0,19	0,80
3	4,42	0,91	0,84	0,18	0,80
4	4,50	0,86	0,84	0,17	0,80
5	5,16	0,80	0,82	0,13	0,77
6	5,92	0,77	0,80	0,10	0,74
7	6,33	0,77	0,79	0,08	0,72

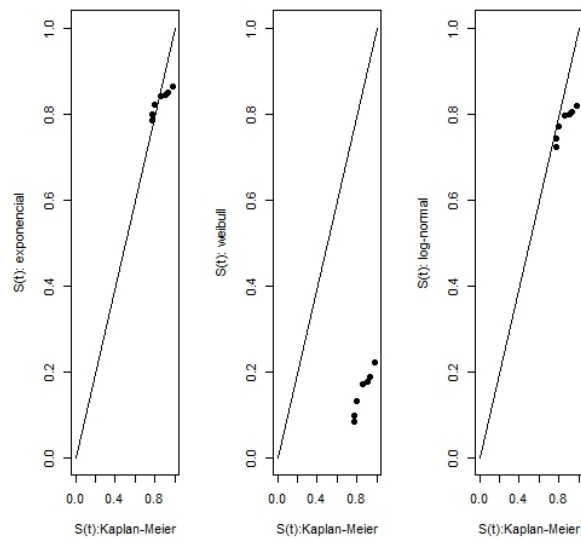


Figura 7: Gráficos dos ajustes das distribuições paramétricas vs Kaplan-Meier

Não houve um ajuste satisfatório das distribuições paramétricas ao tempo até o divórcio (FIGURA 7), apenas a distribuição weibull teve um comportamento um pouco melhor indicando que em uma possível busca por outras distribuições deva-se se levar em conta distribuições assimétricas a direita.

## 4 Conclusão

Por meio das técnicas utilizadas nesse trabalho evidenciou-se que análise de sobrevivência é uma importante ferramenta para modelar o tempo de sobrevivência de casamentos. De forma geral as funções de sobrevivência estimadas por Kaplan-Meier indicam que após 3,5 anos a probabilidade de sobrevivência do casamento diminuí consideravelmente.

# Referências

- ARAÚJO, T. B. L. **Casamento contemporâneo: um olhar clínico sobre os laços conjugais**. Curitiba, 2006.
- BALAKRISHNAN, N. **Handbook of the Logistic Distribution**. New York: Dekker, 1992.
- CARVALHO M., ANDREOZZI V., CODEÇO C., BARBOSA M.T., SHIMUKURA. **Análise de sobrevivência. Teoria e Aplicações em Saúde**. Rio de Janeiro, Editora Fiocruz, 2005.
- COSTA, G.P. **A cena conjugal**. Porto Alegre: Artes Médicas Sul, 2000.
- IBGE. Instituto Brasileiro de Geografia e Estatística. Estatísticas do Registro Civil. 2012, <<http://www.ibge.gov.br/home/estatistica/populacao/registrocivil/2012/default.shtm>>
- FERREIRA, A. M. **Testes não-paramétricos**. Escola Superior Agrária Castelo Branco, 2014. <[http://docentes.esa.ipcb.pt/mede/apontamentos/testes\\_ao\\_parametricos.pdf](http://docentes.esa.ipcb.pt/mede/apontamentos/testes_ao_parametricos.pdf)>. Acesso em 13 de novembro de 2014
- GIOLO, S. R., COLOSIMO, E. A. **Análise de sobrevivência aplicada**. São Paulo, Edgar Blücher, 2006.
- KAPLAN, E. L.; MEIER, P.. **Nonparametric estimation from incomplete observations**, 1958.
- LOUZADA, F., FERREIRA, P. H., DINIZ, C. A. R. On the impact of disproportional samples in credit scoring models: An application to a brazilian bank data. **Expert Systems with Applications**. Volume 39, Issue 9, 2012, p.80718078.
- R Development Core Team (2012). **R: A language and environment for statistical computing**. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 nov. 2014.
- STRAPASSON, E. **Comparação de modelo com censura intervalar**. Tese (Doutorado)-Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2007

TACONELI, J.P. **Modelo de Mistura Paramétrico com Fragilidade na Presença de Covariáveis**, Dissertação (Mestrado) Universidade Federal de São Carlos, p.66, 2013.

WEIBULL, W. A Statistical Representation of Fatigue Failure in Solids. Royal Institute of Technology, 1954.



# APÊNDICE A – Apêndice

Quadro com os dados coletados no 1 Cartório de Campina Grande - PB

Obs	tempo	evento	idade homens	idade mulheres	fator homens	fator mulheres
1	6,33	Censura	22	23	Grupo 1	grupo 2
2	6,33	Censura	52	22	grupo 2	grupo 2
3	6,33	Censura	23	22	Grupo 1	grupo 2
4	6,33	Censura	24	21	Grupo 1	Grupo 1
5	6,33	Censura	23	21	Grupo 1	Grupo 1
6	6,33	Censura	25	29	grupo 2	grupo 2
7	6,33	Censura	22	31	Grupo 1	grupo 2
8	6,33	Censura	24	17	Grupo 1	Grupo 1
9	6,33	Censura	25	18	grupo 2	Grupo 1
10	4,25	Falha	2	16	Grupo 1	Grupo 1
11	6,33	Censura	20	20	Grupo 1	Grupo 1
12	6,33	Censura	27	28	grupo 2	grupo 2
13	6,33	Censura	21	16	Grupo 1	Grupo 1
14	6,33	Censura	24	21	Grupo 1	Grupo 1
15	6,33	Censura	26	26	grupo 2	grupo 2
16	5,16	Falha	20	18	Grupo 1	Grupo 1
17	5,16	Falha	71	59	grupo 2	grupo 2
18	6,33	Censura	26	24	grupo 2	grupo 2
19	4,25	Falha	28	26	grupo 2	grupo 2
20	6,33	Censura	30	24	grupo 2	grupo 2
21	6,33	Censura	25	31	grupo 2	grupo 2
22	6,33	Censura	18	18	Grupo 1	Grupo 1
23	5,92	Falha	25	21	grupo 2	Grupo 1
24	4,42	Falha	23	25	Grupo 1	grupo 2
25	6,33	Censura	23	19	Grupo 1	Grupo 1
26	6,33	Censura	22	19	Grupo 1	Grupo 1
27	6,33	Censura	22	24	Grupo 1	grupo 2
28	6,33	Censura	27	26	grupo 2	grupo 2
29	6,33	Censura	20	18	Grupo 1	Grupo 1
30	6,33	Censura	20	18	Grupo 1	Grupo 1
31	5,16	Falha	22	17	Grupo 1	Grupo 1
32	6,33	Censura	21	20	Grupo 1	Grupo 1
33	6,33	Censura	29	22	grupo 2	grupo 2
34	6,33	Censura	25	18	grupo 2	Grupo 1
35	6,33	Censura	26	17	grupo 2	Grupo 1
36	6,33	Censura	29	36	grupo 2	grupo 2
37	6,33	Censura	25	17	Grupo 1	Grupo 1
38	4,50	Falha	24	16	grupo 2	Grupo 1
39	6,33	Censura	28	21	grupo 2	Grupo 1
40	6,33	Censura	27	22	Grupo 1	grupo 2
41	6,33	Censura	20	17	grupo 2	Grupo 1
42	3,83	Falha	25	22	grupo 2	grupo 2
43	6,33	Censura	24	31	Grupo 1	grupo 2
44	4,5	Falha	24	16	Grupo 1	Grupo 1

## ANEXO A – Anexo

```

dados=read.table("dados_erivaldo.txt",h=T)
attach(dados)
head(dados)
require(survival)
ekm=survfit(Surv(tempo,evento)~fator2, conf.type = "log")
summary(ekm)
survdiff(Surv(tempo,evento)~fator2,rho=0)
plot(ekm,lty=c(1,4),mark.time=F,xlab="tempo até o divórcio(anos)",
ylab="S(t)",ylim=c(0.70,1.00), xlim=c(3.5,6.5))
plot(ekm, fun = function(x) 1-x, xlab="tempo até o divórcio(anos)",
ylab="h(t)")
ekm1=survfit(Surv(tempo,evento)~1, conf.type = "log")
plot(ekm1,lty=c(1,4),mark.time=F,xlab="tempo até o divórcio(anos)",
ylab="S(t)",ylim=c(0.70,1.00), xlim=c(3.5,6.5))
#####
ekm=survfit(Surv(tempo,evento)~fator1, conf.type = "log")
summary(ekm)
survdiff(Surv(tempo,evento)~fator1,rho=0)
plot(ekm,lty=c(1,4),mark.time=F,xlab="tempo até o divórcio(anos)",
ylab="S(t)",ylim=c(0.70,1.00), xlim=c(3.5,6.5))
plot(ekm, fun = function(x) 1-x, xlab="tempo até o divórcio(anos)",
ylab="h(t)")
#####
ekm=survfit(Surv(tempo,evento)~fator1+fator2, conf.type = "log")
summary(ekm)
survdiff(Surv(tempo,evento)~fator1+fator2,rho=0)
plot(ekm,lty=c(1,4),mark.time=F,xlab="tempo até o divórcio(anos)",
ylab="S(t)",ylim=c(0.70,1.00), xlim=c(3.5,6.5))
plot(ekm, fun = function(x) 1-x, xlab="tempo até o divórcio(anos)",
ylab="h(t)")

```

```
#####
adjust=survreg(Surv(tempo,evento)~1,dist='exponential')
adjust
alpha=exp(adjust$coefficients[1])
alpha
adjust1=survreg(Surv(tempo,evento)~1,dist='weibull')
adjust1
alpha1=exp(adjust1$coefficients[1])
alpha1
gama=1/adjust1$scale
cbind(gama,alpha1)
adjust2=survreg(Surv(tempo,evento)~1,dist='lognorm')
adjust2
adjust3=survreg(Surv(tempo,evento)~1,dist='loglogistic')
adjust3

d=(tempo/exp(2.141687))^(1/0.2549584)
stll=1/(1+d)
stll

gama=1/adjust1$scale

ekm1=survfit(Surv(tempo,evento)~1)
time=ekm1$time
st=ekm1$surv
ste=exp(-time/alpha)
stw=exp(-time/alpha1)^gama
stl=pnorm((-log(time)+2.16)^0.4506)
stll=1/(1+d)

require(xtable)
xtable(cbind(time,st,ste,stw,stl,stll))

par(mfrow=c(1,3))
plot(st,ste,pch=16,ylim=range(c(0.0,1)),xlim=range(c(0,1)),
```

```

xlab="S(t):Kaplan-Meier",ylab="S(t): exponencial")
lines(c(0,1),c(0,1), type="l",lty=1)

par(mfrow=c(1,1))
plot(st,stw,pch=16,ylim=range(c(0.0,1)),xlim=range(c(0,1)),
xlab="S(t):Kaplan-Meier",ylab="S(t): weibull")
lines(c(0,1),c(0,1), type="l",lty=1)

par(mfrow=c(1,1))
plot(st,stl,pch=16,ylim=range(c(0.0,1)),xlim=range(c(0,1)),
xlab="S(t):Kaplan-Meier",ylab="S(t): log-normal")
lines(c(0,1),c(0,1), type="l",lty=1)

###logistica

#####Gráficos Linearizados #####

invst=qnorm(stl)
par(mfrow=c(1,3))
plot(time, -log(stl), pch=16,xlab="tempos",ylab="-log(S(t))")
plot(log(time), log(-log(stl)), pch=16,xlab="log(tempos)",
ylab="log(-log(S(t)))")
plot(log(time),invst, pch=16,xlab="log(tempos)",
ylab=expression(Phi^-1*(S(t))))

##### Teste da razão de verossimilhança #####

adjust$loglik[2]
adjust1$loglik[2]
adjust2$loglik[2]

```