

MARIA NILSA MARTINS DE ARAUJO

ANÁLISE DE SOBREVIVÊNCIA DO TOMATEIRO A *Phytophthora infestans*

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Estatística Aplicada
e Biometria, para obtenção do título de
Magister Scientiae.

VIÇOSA
MINAS GERAIS - BRASIL
2008

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

A663a
2008

Araújo, Maria Nilsa Martins, 1961-
Análise de sobrevivência do tomateiro a
Phytophthora infestans / Maria Nilsa Martins Araújo.
– Viçosa, MG, 2008.
xi, 53f. : il. (algumas col.) ; 29cm.

Orientador: Sebastião Martins Filho.
Dissertação (mestrado) - Universidade Federal de
Viçosa.
Referências bibliográficas: f. 49-53.

1. Técnica de Kaplan-Meier. 2. Distribuição (Teoria da
probabilidade) - Weibull. 3. Distribuição (Teoria da
probabilidade) - Log-normal. 4. Famílias exponenciais
(Estatística). 5. Tomate - Doenças e pragas.
6. *Phytophthora infestans*. 7. Sobrevivência.
I. Universidade Federal de Viçosa. II. Título.


CDD 22.ed. 519.2


MARIA NILSA MARTINS DE ARAUJO

ANÁLISE DE SOBREVIVÊNCIA DO TOMATEIRO A *Phytophthora infestans*

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de
Pós-Graduação em Estatística Aplicada
e Biometria, para obtenção do título de
Magister Scientiae.


APROVADA: 05 de setembro de 2008.


Prof. Fabyano Fonseca e Silva
(Co-Orientador)


Prof. Paulo Roberto Cecon
(Co-Orientador)


Prof. Antônio Policarpo Souza Carneiro


Pesq. Luciano Viana Cota


Prof. Sebastião Martins Filho
(Orientador)

“Nas tendas dos justos há jubiloso cântico de vitória; a destra do Senhor faz proezas.” (Sl. 118:150)

A DEUS dedico esta vitória.

Aos meus queridos pais José e Amorosa, fonte de confiança, apoio e luta na educação dos filhos.

Ao meu esposo Lúcio, cujo apoio, compreensão e incentivo foram importantes nesta etapa da minha vida profissional.

Aos meus queridos filhos Lucas e Thiago, presentes de Deus que enchem a minha vida de amor e alegria.

Aos meus irmãos e suas famílias Manoel e Naide; Suely; Elton e Roca; Tony; Orly e João Helton; e Sandra, fortalezas em todos os momentos da minha vida, especialmente durante meus estudos do Mestrado.

OFEREÇO

AGRADECIMENTOS

A Deus e a Nossa Senhora, pelo dom da vida, cuja fé neles me deu a força necessária para vencer todos os desafios.

À Universidade Federal de Viçosa (UFV) e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria (PPESTBIO), pela oportunidade de realização de trabalhos em minha área de pesquisa.

À Escola Agrotécnica Federal de Salinas (EAFSalinas), Salinas, MG, pela oportunidade concedida para a realização do Curso de Mestrado em Estatística, em nome do Professor Adalcino França Júnior.

Ao Professor Sebastião Martins Filho, pela amizade, orientação e pelo apoio na execução deste trabalho.

Ao Professor Fabiano Fonseca e Silva, pela co-orientação, ajuda e pelo apoio na execução e correção da dissertação e das análises estatísticas, pelo ensinamento, estímulo, exemplo e pela amizade. Ser-lhe-ei eternamente grata.

Ao Professor Paulo Roberto Cecon, pela amizade, atenção, ajuda e pelo incentivo na realização e correção deste estudo.

Aos meus Professores Sebastião, Antônio Policarpo, Cosme, José Ivo, Cecon, Carlos Henrique e Rita Márcia, pelos conhecimentos transmitidos e pela paciência.

Aos meus queridos pais José Antônio de Araújo e Amorosa Martins de Araújo, pelo incentivo, apoio e por não medirem esforços para a minha educação.

Ao meu esposo Lúcio, pela compreensão e apoio em todos os momentos desta importante etapa de minha vida profissional, e à sua família, pela compreensão nestes anos de ausência.

Ao meu cunhado João Helton e à minha irmã Orly, pelo carinho e ajuda, pelas palavras amigas nos momentos difíceis e pela constante torcida pelo meu sucesso.

Aos meus irmãos Elton e Sandra, pelo apoio e ajuda nos momentos em que mais precisei, pelo amor e carinho de sempre.

Ao meu amigo e colega de trabalho Professor Murilo Nonato Bastos, pela força e pela torcida pelo meu sucesso.

À minha amiga dona Beta e seus filhos, pela fé, pelas orações e pelo constante incentivo.

Ao casal Flávio e Dorinha, pelo afetuoso acolhimento em Viçosa.

Aos meus colegas do PPESTBIO, em especial a Alex, Danilo, Thiago, Andréia, Telma, Willerson, Ana Carolina, Moisés e Flávia, pelos bons momentos, pela união nos estudos, pelas trocas de experiências, pelo estímulo e pela amizade.

Aos membros da minha Banca Examinadora Paulo Roberto Cecon, Antônio Policarpo Sousa Carneiro e Luciano Viana Costa, pelas colaborações e valiosas sugestões, o que muito enriqueceu esta pesquisa.

À CAPES, pelo fundamental suporte financeiro.

Ao secretário de Pós-Graduação Altino, pela gentileza, presteza e amizade nestes anos de curso.

Aos meus colegas e amigos da EAFSalinas Geraldo Magela, Soraya, Ivoneide, Daniela, Edilene, Charles, Gessionei, Luiz, Ramon, Máximo, Roberto Guimarães e Gilmar, pelo apoio.

Aos meus sobrinhos Diogo, Klenner, Henrique, Douglas, Anny, Nádilla, Binha, Bruna e Maria Eduarda, bem como aos meus afilhados Mayra e Breno, por tornarem minha vida bonita e alegre e, apesar da distância física, terem caminhado comigo nesta jornada.

A todos que, de alguma maneira, contribuíram para que eu conseguisse chegar até aqui, seja com um sorriso, seja com um incentivo ou simplesmente com a força positiva do pensamento em mim – agradeço a todos, com muito carinho.

BIOGRAFIA

MARIA NILSA MARTINS DE ARAUJO, filha de José Antônio de Araújo e Amorosa Martins de Araújo, nasceu no dia 20 de agosto de 1961, em Salinas, MG.

Em 1990, ingressou no Curso de Matemática da Universidade Estadual de Montes Claros – UNIMONTES, em Montes Claros, MG, graduando-se em dezembro de 1994.

Entre 1995 e 1996, foi designada para exercer o cargo de Professora de Ensino de 1º e 2º Graus na Escola Estadual Dr. Oswaldo Prediliano Santana e Escola Estadual Coronel Idalino Ribeiro, em Salinas, MG.

Em 1996, ingressou no Curso de Especialização em Metodologia do Ensino da Matemática das Faculdades Clarentianas, Batatais, SP, concluindo-o nesse mesmo ano.

Em 1996, foi contratada como Professora Substituta de Ensino de 1º e 2º Graus pela Escola Agrotécnica Federal de Salinas.

Em dezembro de 1997, foi nomeada no cargo de Professor de Ensino de 1º e 2º Graus na Escola Agrotécnica Federal de Salinas.

De 2002 a 2004, foi representante do Corpo Docente do Conselho-Diretor da Escola Agrotécnica de Salinas e representante da SEMTEC no Conselho-Diretor da Escola Agrotécnica Federal de Guanambi, BA.

Em abril de 2002, foi designada para a Função Gratificada, código FG-04, de Supervisora Pedagógica, cargo que exerceu até 2006.

Em outubro de 2006, ingressou no Programa de Pós-Graduação, em nível de Mestrado, em Estatística Aplicada e Biometria da Universidade Federal de Viçosa, em Viçosa, MG, submetendo-se à defesa da dissertação em setembro de 2008.

SUMÁRIO

	Página
RESUMO.....	viii
ABSTRACT.....	x
1. INTRODUÇÃO.....	01
2. REVISÃO DE LITERATURA.....	04
2.1. Análise de sobrevivência.....	04
2.2. Fitopatologia e curvas de progresso de doenças.....	07
2.3. Fitopatologia e técnicas de análise de sobrevivência.....	09
2.4. Importância do estudo da requeima do tomateiro.....	10
3. MATERIAL E MÉTODOS.....	12
3.1. Descrição dos dados.....	12
3.2. Ajuste de curvas de progresso da doença e agrupamento dos acessos.....	13
3.3. Tempo até a severidade de 5% da doença (regressão inversa).....	14
3.4. Estimação de parâmetros na análise de sobrevivência.....	15
3.4.1. Técnicas não-paramétricas utilizadas na análise de sobrevivência.....	15
3.4.1.1. Estimador de Kaplan-Meier.....	15
3.4.1.2. Teste <i>Logrank</i> para comparação de curvas de sobrevivência.....	18
3.4.2. Modelos probabilísticos utilizados na análise de sobrevivência.....	21

3.4.2.1. Modelo exponencial.....	22
3.4.2.2. Modelo Weibull.....	23
3.4.2.3. Modelo Log-normal.....	24
3.4.2.4. Modelo Gama generalizado.....	25
3.4.3. Estimação de parâmetros dos modelos probabilísticos.....	26
3.4.4. Escolha do modelo probabilístico.....	27
3.4.5. Métodos para comparação dos parâmetros dos modelos probabilísticos.....	28
3.5. Softwares utilizados para a análise de sobrevivência.....	29
4. RESULTADOS E DISCUSSÃO.....	30
4.1. Ajuste de curvas de progresso da doença e agrupamento dos acessos.....	30
4.2. Tempo até a incidência de 5% da doença (regressão inversa).....	33
4.3. Análise de sobrevivência.....	35
4.3.1. Estimador de Kaplan-Meier e Teste <i>Logrank</i>	36
4.3.2. Modelos probabilísticos.....	39
4.3.3. Comparação dos parâmetros dos modelos probabilísticos.....	43
5. CONCLUSÕES.....	47
6. REFERÊNCIAS.....	49

RESUMO

ARAUJO, Maria Nilsa Martins de, M. Sc., Universidade Federal de Viçosa, setembro de 2008. **Análise de sobrevivência do tomateiro a *Phytophthora infestans***. Orientador: Sebastião Martins Filho. Co-Orientadores: Fabyano Fonseca e Silva e Paulo Roberto Cecon.

A requeima causada por *Phytophthora infestans* caracteriza-se por ser uma doença agressiva e de grande impacto destrutivo, podendo limitar ou até mesmo impedir o cultivo econômico do tomateiro sob condições de alta umidade e baixas temperaturas. Diante dos problemas que a requeima pode provocar às lavouras de tomate, este trabalho teve por objetivos: 1) ajustar modelos para descrever o progresso da doença e formar grupos de acessos de tomateiro com curvas semelhantes; 2) estimar dados referentes ao número de dias até atingir 5% de severidade da doença, por meio de regressão inversa; 3) ajustar curvas de sobrevivência por meio do estimador de Kaplan-Meier para grupos de acessos e compará-las mediante o uso do teste *Logrank*; 4) ajustar curvas de sobrevivência por meio de modelos probabilísticos e compará-las com a técnica não-paramétrica de Kaplan-Meier. Utilizando dados reais sobre a requeima do tomateiro, foi possível ajustar o modelo exponencial ($Y = y_0 \exp(rX)$) para descrever o progresso da doença. As médias das estimativas dos parâmetros foram submetidas à análise de agrupamento pelo método Centróide, o que gerou 10 grupos de acessos, sendo o tempo até a incidência de 5% da doença calculado via regressão inversa. Foram utilizadas técnicas não-paramétricas para estimar a função de sobrevivência por meio do estimador de Kaplan-Meier e para comparar as curvas de sobrevivência pelo teste *Logrank*. Foi também ajustada a função de sobrevivência, empregando-se os modelos

probabilísticos Exponencial, Weibull e Log-normal, os quais foram comparados por meio do Teste da Razão da Verossimilhança (TRV), considerando-se o modelo Gama generalizado por ser caso geral para esses modelos. A metodologia utilizada permitiu ajustar o modelo Exponencial para descrever o progresso da requeima do tomateiro e agrupar os acessos estudados em 10 grupos. O acesso BGH-6 sofreu um progresso de doença menor que os demais, caracterizando-se, assim, sua maior resistência à enfermidade. A regressão inversa possibilitou estimar o tempo até a ocorrência de 5% da severidade da requeima do tomateiro. Pela técnica não-paramétrica de Kaplan-Meier, foi possível estimar as curvas de sobrevivência dos acessos de tomateiro pertencentes aos grupos 1, 2, 4, 6 e 8. Utilizando o teste *Logrank*, pode-se concluir que a maioria das comparações duas a duas foi significativa ($p < 0,05$), exceto nas comparações dos grupos 2x4, 4x8 e 6x8. O uso dos modelos probabilísticos Exponencial, Weibull e Log-normal possibilitou a estimação das curvas de sobrevivência nos grupos 2, 4, 6 e 8, exceto no grupo 4, em que o modelo Weibull não foi adequado. Comparando os modelos probabilísticos com a técnica não-paramétrica, as curvas dos modelos probabilísticos dos grupos 2 e 4 apresentaram ajustes satisfatórios com relação à curva estimada por Kaplan-Meier.

ABSTRACT

ARAÚJO, Maria Nilsa Martins de, M. Sc., Universidade Federal de Viçosa, September, 2008. **Analysis of the survival of the tomato plant *Phytophthora infestans*.** Adviser: Sebastião Martins Filho. Co-Advisers: Fabyano Fonseca e Silva and Paulo Roberto Cecon.

Reburning caused by *Phytophthora infestans* is characterized as an aggressive disease of great destructive impact, capable of limiting or even hindering the economic cultivation of the tomato plant under conditions of high humidity and low temperatures. In view of the problems reburning can cause to tomato plant crops, this work aimed to: 1) fit models to describe the progress of the disease and form groups of tomato accesses with similar curves; 2) estimate data referring to the number of days to reach 5% severity of the disease, by means of inverse regression; 3) fit survival curves by means of the Kaplan-Meier estimator for the access groups and compare them by means of the *Logrank* test; 4) fit survival curves by means of probabilistic models and compare these curves with Kaplan Meir's non-parametric technique. Using tomato reburning real data, it was possible to fit the exponential model ($Y = y_0 \exp(rX)$) to describe the disease's progress. The means of the parameter estimates were submitted to grouping analysis using the centroid method, generating 10 access groups. Time up to 5% of the disease was calculated via inverse regression. Non-parametric techniques were used to estimate survival function by means of the Kaplan-Meier's estimator to compare the survival curves by the *Logrank* test. The survival function was also fit using the probabilistic models, exponential Weibull and Log-normal, respectively, which were compared by means of the verisimilitude ratio test (VRT), considering the

generalized Gamma model, as a general case for these models. The methodology applied allowed fitting the exponential model to describe tomato plant reburning progress and to regroup the accesses studied in the 10 groups. The access BGH-6 obtained a smaller disease progress than the others, thus characterizing its higher resistance to the disease; An inverse regression allowed time estimation up to the occurrence of 5% of the severity of the tomato plant reburning. The Kaplan-Meier 's non-parametric technique allowed estimating the survival curves of the tomato plant accesses belonging to the groups 1, 2, 4, 6 and 8. Utilizing the *Logrank* test, it could be concluded that most two-by-two comparisons were significant ($p < 0.05$), except in the comparisons of groups 2x4, 4x8 and 6x8. The use of the probabilistic models, exponential Weibull and Log-normal allowed estimating the survival curves of groups 2, 4, 6 and 8, except for group 4, to which the Weibull model was not adequate. Comparing the probabilistic models with the non-parametric technique, the curves of the probabilistic models of groups 2 and 4 presented satisfactory results, compared to the curve estimated by Kaplan-Meier.

1. INTRODUÇÃO

O tomateiro (*Lycopersicon esculentum* Mill.) industrial é considerado uma das culturas de grande importância para as Regiões Centro-Oeste e Sudeste, devido à sua utilização na indústria de extrato de polpa (CAMARGO FILHO; MAZZEI, 1997; SILVA; GIORDANO, 2000). Entretanto, inúmeros fatores vêm limitando a produção de tomate industrial no Brasil, sendo as doenças um dos mais importantes, por reduzirem a produtividade e a qualidade dos frutos. Entre as doenças que atacam o tomateiro industrial, a requeima causada por *Phytophthora infestans* (Mont.) De Bary é a mais destrutiva e pode comprometer todo o campo de produção em poucos dias (STEVENSON, 1983; LOPES; SANTOS, 1994). Os sintomas da requeima ocorrem nas folhas, hastes, nos frutos e pecíolos do tomateiro, com aspecto semelhante ao da queima (KUROZAWA; PAVAN, 1997).

A análise de sobrevivência é usada quando a variável resposta corresponde ao período de tempo até a ocorrência de algum evento de interesse. Esse período também é definido como tempo de falha, sendo o termo falha a ocorrência do evento em questão. Uma característica importante dos dados de sobrevivência é a presença de censuras, as quais são consideradas observações incompletas e se caracterizam como dados de indivíduos em que a ocorrência do evento não foi verificada.

A censura é o fator que diferencia a análise de sobrevivência de outras técnicas de análise de dados. Entre os métodos estatísticos considerados em análise de sobrevivência, um dos mais utilizados é a estimação de curvas de sobrevivência, as quais relacionam a probabilidade de o evento de interesse ocorrer em função dos valores assumidos pela variável em questão, ou seja, os tempos observados. Após essa

estimação, também é de grande interesse a comparação entre as curvas provenientes de diferentes tratamentos.

Nas Ciências Agrárias, a análise de sobrevivência tem grande número de aplicações, por exemplo, no estudo de germinação de sementes ou no estudo do controle de plantas daninhas; no tempo de vida de insetos após a aplicação de determinado tratamento; principalmente em Fitopatologia, para analisar o tempo até aparecer os primeiros sintomas de uma doença, na germinação de esporos de um fungo; e na ocorrência do desfolhamento causado por uma doença foliar (SCHERM; OJIAMBO, 2004).

Ainda, na área da Fitopatologia, tal procedimento estatístico também pode ser utilizado, a fim de identificar tratamentos (diferentes cultivares, diferentes sistemas de manejo etc.) que fornecem maior sobrevivência em relação a doenças que ocorrem em uma cultura de interesse. Porém, devido a aspectos práticos de coleta de dados, muitas vezes a variável tempo até a doença atingir determinado nível de incidência (variável de interesse), é de difícil mensuração. Essa dificuldade ocorre porque, ao se realizar uma avaliação, a doença já pode ter ultrapassado o nível de incidência fixado pelo pesquisador ou, ainda, não chegou a atingi-lo.

Uma forma prática e eficiente de obter esses dados é ajustar modelos de regressão que descrevem a incidência (severidade da doença) em função do tempo, os quais são denominados modelos de progresso de doenças e comumente utilizados em Fitopatologia. Depois desse, é possível adotar um nível de interesse, e por regressão inversa (fixar valor da variável dependente e obter o valor esperado para a variável independente), obter o tempo até a ocorrência do nível desejado.

Mediante um conjunto de dados com observações referentes ao tempo até a ocorrência de determinado nível de incidência, e os respectivos tratamentos e repetições, é possível estimar curvas de sobrevivência para cada tratamento e, também, compará-las, a fim de identificar aquelas que apresentam maior probabilidade de sobrevivência.

Utilizando dados reais sobre a requeima do tomateiro causada por *Phytophthora infestans*, (Mont.) De Bary, o objetivo deste trabalho foi: 1) ajustar modelos de regressão para descrever o progresso da doença e formar grupos de acessos de tomateiro para identificar aqueles com maior sobrevivência; 2) estimar dados referentes ao tempo até atingir 5% de severidade da doença, por meio de regressão inversa; 3) ajustar curvas de sobrevivência por meio do estimador de Kaplan-Meier,

para grupos de acessos, e compará-las mediante o uso do teste *Logrank*; 4) ajustar curvas de sobrevivência por meio de modelos probabilísticos e compará-las com a técnica não-paramétrica, de Kaplan-Meier.

2. REVISÃO DE LITERATURA

2.1. Análise de sobrevivência

A Análise de Sobrevivência é um conjunto de técnicas e modelos estatísticos usados na análise de experimentos, cuja variável resposta é o tempo até a ocorrência de um evento de interesse. Os indivíduos sob estudo podem ser animais, seres humanos, plantas, equipamentos etc. No entanto, o evento de interesse pode ser: morte, remissão de uma doença, reação de um medicamento, quebra de um equipamento eletrônico, queima de uma lâmpada etc. A principal característica dos dados de sobrevivência é a presença de censuras, que é a observação parcial da resposta. Essa informação, apesar de incompleta, é útil e importante para a análise (AALEN, 1987).

Em engenharia, são comuns os estudos em que produtos ou componentes são colocados sob teste para se estimar em características relacionadas aos seus tempos de vida, como o tempo médio ou a probabilidade de certo produto durar mais do que cinco anos. Os engenheiros denominam essa área de Confiabilidade, mas as técnicas são as mesmas da Análise de Sobrevivência (COLOSIMO; GIOLO, 2006). Exemplos nessa área podem ser encontrados em Freitas e Colosimo (1997).

A análise de sobrevivência é o ramo da estatística que trata da análise de dados provenientes de variáveis que assumem valores positivos (CHALITA et al., 2006). O que diferencia a análise de sobrevivência de outras análises é a presença de censura, que torna a observação incompleta. Isso porque, na maioria das vezes, se está interessado em observar o tempo até a falha de algum evento. Quando a falha não acontece, diz-se que a observação foi censurada, e é preciso inserir tal informação na análise. Segundo

Klein (1997), diante desse tipo de dados, o interesse está em usar técnicas estatísticas que permitam a inserção dessas observações. Esse mesmo autor mencionou que uma das mais importantes dessas técnicas é a estimação da função de sobrevivência, que é dada pela probabilidade de que um item falhe até determinado tempo t .

Na análise de sobrevivência, os parâmetros mais usados são a probabilidade de sobrevivência no curso de cada um dos intervalos considerados e a probabilidade de sobrevivência acumulada, isto é, a probabilidade de sobreviver do tempo zero até o tempo final considerado (TEIXEIRA et al., 2002). Esta última equivale à probabilidade de sobreviver em todos os intervalos anteriores ao momento considerado e, usualmente, é denominada função de sobrevivência $S(t)$. Outra função importante em análise de sobrevivência é a função de taxa de falha ou de risco $\lambda(t)$ que descreve a forma em que a taxa instantânea de falha muda com o tempo.

As funções de taxa de falha podem ser crescente, constante ou decrescente (Figura 1). A função crescente indica que a taxa de falha de um indivíduo aumenta com o transcorrer do tempo. Esse comportamento indica um efeito gradual de envelhecimento. A função constante indica que a taxa de falha não se altera com o passar do tempo. A função decrescente evidencia que a taxa de falha diminui à medida que o tempo passa.

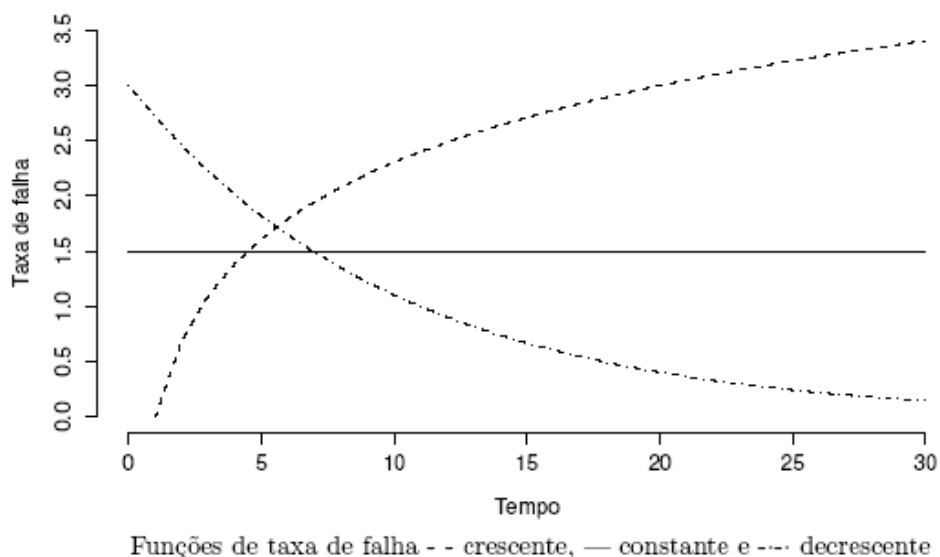


Figura 1 - Representação gráfica de três funções de taxa de falhas.

A estimativa da probabilidade de sobrevida é, com certeza, mais válida e mais precisa no período inicial do seguimento, no qual estão disponíveis informações sobre a maioria dos pacientes. Nos períodos posteriores, as informações podem ficar limitadas devido às perdas de seguimento e ao pequeno número de eventos (FLETCHER et al., 1996).

Somente nas décadas de 1950 e de 1960 apareceram as primeiras propostas de estimadores das probabilidades de sobrevida, $S(t)$, que incorporavam a censura, ou seja, modelos para observações incompletas. Os métodos de estimação da função de sobrevivência são o atuarial (CHIANG, 1968), quando os dados estão em intervalos, e o de Kaplan-Meier (KAPLAN; MEIER, 1958), quando os dados são o tempo exato de sobrevivência.

O método atuarial e o método de Kaplan-Meier, assumiram, inicialmente, que as observações censuradas têm a mesma probabilidade de sofrerem o evento que aquelas que permanecem em observação. A distinção essencial entre eles é que este último elimina a necessidade de assumir que as censuras das observações ocorram uniformemente durante esse intervalo. Assume-se apenas que as observações censuradas teriam a mesma experiência futura do que aquelas que continuam sendo observadas (KAHN; SEMPOS, 1989). O método de Kaplan-Meier pode ser utilizado para qualquer tamanho de amostra em estudo, mas é especialmente útil naqueles estudos com um número pequeno de observações, enquanto o método atuarial é mais apropriado para grandes amostras (LEE, 1992).

Após a estimação das curvas de sobrevivência, fundamentadas na probabilidade citada anteriormente ($S(t)$), caso os dados sejam provenientes de vários tratamentos o interesse será comparar esses tratamentos para saber quais deles fornecem maior probabilidade de sobrevivência. Para isso, existem vários métodos, sendo os não-paramétricos mais utilizados por não necessitarem de pressuposições (ALLISON, 1995). Entre esses, destacam-se o *Logrank* (MANTEL, 1966) e o Wilcoxon (GEHAN, 1965).

O teste *Logrank* considera uma seqüência de tabelas de contingência baseadas no número de indivíduos em risco (SILVEIRA; CORRENTE, 2004). Já o teste de Wilcoxon considera uma estatística ponderada pelo número de indivíduos em risco em cada tempo, o que favorece os indivíduos no início do estudo, já que o número de indivíduos decresce com o passar do tempo. Esses mesmos autores também propuseram o teste de Friedman, o qual, segundo eles, apresenta mais diferença significativa entre os

tratamentos aplicados aos dados, enquanto os testes *Logrank* e Wilcoxon não apresentaram tal significância, sendo, então, mais rigorosos. Porém, esses dois testes apresentam bons resultados em relação a seu poder, e nada se tem na literatura a respeito do poder do teste de Friedman, o qual ainda necessita de estudos de simulação para verificar a sua real aplicabilidade a dados de sobrevivência.

2.2. Fitopatologia e curvas de progresso de doenças

Fitopatologia é uma palavra de origem grega (Phyton = planta, Pathos = doença e Logos = estudo) e indica a ciência que estuda as doenças das plantas, abrangendo todos os seus aspectos, desde a diagnose, sintomatologia, etiologia, epidemiologia, até o seu controle. O desenvolvimento da Fitopatologia como ciência data de período relativamente curto. Entretanto, o relato de doenças em plantas é bastante antigo e, desde que o homem passou a fixar-se e desenvolver a agricultura como forma de obter alimentos para sua sobrevivência, passou também a enfrentar problemas relacionados a perdas completas de plantações por questões de doenças e pragas. Na Bíblia são encontradas algumas das referências mais antigas com relação a doenças de plantas, a exemplo da ferrugem dos cereais e doenças em videiras e olivais. Esses fenômenos eram normalmente atribuídos a causas místicas e, ou, castigos divinos.

Atualmente, o crescente e contínuo aumento na população mundial fez crescer a preocupação com a quantidade e qualidade dos alimentos produzidos. Uma projeção levar a imaginar que, no futuro, não haverá alimentos suficientes para alimentar a população terrestre. Dessa forma, torna-se necessário buscar instrumentos que permitam aumento na produção de alimentos. Entretanto, o desenvolvimento da agricultura esta diretamente relacionada com algumas questões básicas, como a influência de solos, clima, pragas e doenças, além, é claro, do desenvolvimento de avanços tecnológicos.

Inserida nesse contexto, a Fitopatologia apresenta-se como ferramenta crucial para subsidiar o desenvolvimento da produção agrícola, buscando solucionar os problemas relacionados com o aparecimento de doenças que reduzem a quantidade e qualidade dos alimentos produzidos.

Dentro da área de Fitopatologia, o progresso de doenças representa a integração dos efeitos do hospedeiro, patógeno e ambiente durante a epidemia e pode oferecer subsídios para analisar, comparar e entender a dinâmica das doenças de plantas (CAMPBELL; MADDEN, 1990). A representação de uma epidemia pode ser feita por

meio da curva de progresso da doença, expressa na plotagem da proporção de doença *versus* tempo. Dessa forma, pode-se determinar a época de início da epidemia, a quantidade de inóculo inicial (Y_0), a taxa de progresso da doença (r), a área abaixo da curva de progresso, a quantidade máxima (Y_{max}) da doença e a duração da epidemia (BERGAMIN FILHO, 1995).

O ajuste de modelos de progresso de doenças geralmente é analisado mediante a análise de regressão não-linear, considerando-se, modelos empíricos com grande aplicação prática (LUTHIANI et al., 2006). Entre esses se destacam o Logístico, Monomolecular, Gompertz e Exponencial (CAMPBELL; MADDEN, 1990). Para a escolha do melhor modelo geralmente levam em consideração o coeficiente de determinação ajustado da análise de regressão (R^2_a), o valor do quadrado médio dos desvios e o gráfico dos resíduos (LUTHIANI et al., 2006).

Uma forma interessante de obter informações a respeito do tempo até a ocorrência de determinado grau de infecção por uma doença em uma cultura é por meio de valores estimados por regressão inversa, considerando-se a curva de progresso da doença. Assim, ajusta-se um modelo de regressão para descrever a severidade em função do tempo; por exemplo, pode-se usar um modelo bem simples como o exponencial:

$$y = y_0 \exp(rx)$$

em que:

y = grau de severidade da doença;

y_0 = quantidade inicial de doença; e

r = taxa de progresso desta doença.

Procedimento análogo foi utilizado por Pires e Colosimo (2005), os quais obtiveram o tempo até crianças atingirem determinados pesos, por meio de modelos não-lineares de curvas de crescimento, de forma que, posteriormente, tais tempos foram considerados como variáveis dependentes em estudos de análise de sobrevivência.

Para simplificar o entendimento dessa metodologia, é possível considerar o exemplo a seguir. Supondo que o modelo estimado seja dado por $y = 3 \exp(0,025x)$ para obter o número de dias, ou seja, um valor para x , até a ocorrência de determinado grau de severidade, por exemplo 10%, pode-se aplicar a regressão inversa da seguinte forma: $10 = 3 \exp(0,025x)$, resultando em $\ln(10/3) = 0,025x$, o que produz um $x = 48,15$ dias. Dessa forma, essa variável x pode ser entendida como uma variável que representa o

tempo até a doença atingir 10% de incidência, portanto passível de ser analisada mediante técnicas de análise de sobrevivência.

Os experimentos conduzidos na área de Fitopatologia, com o intuito de estudar curvas de progresso de doenças, geralmente apresentam t tratamentos e r repetições, de forma que as tr unidades experimentais são avaliadas no decorrer do tempo (CAMPBELL; MADDEN, 1990). Para utilizar a regressão inversa apresentada anteriormente, deve-se, então, ajustar o modelo que descreve o progresso da doença em função do tempo, para cada unidade experimental, o que resulta em tr valores de tempos até a ocorrência de certo grau de doença. Assim, mediante um conjunto de dados, com tratamentos e repetições, isso permite usar técnicas de análise de sobrevivência para comparar os diferentes tratamentos em relação às suas eficiências no controle de doenças de plantas. Tais eficiências são avaliadas por meio da comparação das curvas de sobrevivência desses tratamentos.

2.3. Fitopatologia e técnicas de análise de sobrevivência

Dados sobre o tempo até a ocorrência de eventos são coletados rotineiramente, tanto no campo quando em laboratórios, na área de epidemiologia de doenças das plantas. Os exemplos incluem o tempo de aparecimento dos primeiros sintomas em uma população de plantas; o tempo de germinação em uma população de esporos fúngicos e o tempo de desfolha em uma população contaminada por determinados patógenos. Apesar da ocorrência comum desses dados em patologia vegetal, a análise de sobrevivência raramente é usada (SCHERM; OJIAMBO, 2004).

Há várias questões que podem ser levantadas por pesquisadores da área de Fitopatologia em relação aos dados de sobrevivência (SCHERM; OJIAMBO, 2004). Em primeiro lugar, pode ser de interesse estimar a distribuição do tempo de sobrevivência para um grupo de indivíduos, ou seja, estimar a curva de sobrevivência desse grupo. Um segundo objetivo pode estar na comparação dessas curvas entre diferentes grupos, os quais caracterizam os tratamentos de interesse para o fitopatologista. Por último, também pode ser de interesse quantificar os efeitos de uma ou mais variáveis independentes (co-variáveis) sobre a sobrevivência, a exemplo da temperatura e umidade, entre outras.

Pereira et al. (2005), estudando o comportamento alimentar de cigarrinhas vetoras de *Xylella fastidiosa*, agente causal da clorose variegada dos citros, avaliaram a

influência das condições hídricas do solo e da temperatura ambiente na sobrevivência do vetor, e para tanto usaram os testes *Logrank* e *Wilcoxon* para comparar as curvas de sobrevivência referentes a cada ambiente considerado.

Em um trabalho inovador, Paula (2007) avaliou a ação de fungos entopatogênicos, geralmente utilizados no controle biológico de insetos, sobre curvas de sobrevivência de mosquitos da dengue. Para tanto, esse autor utilizou o estimador de Kaplan-Meier para obter as curvas, e utilizou o teste *Logrank* para compará-las. Neste estudo, primeiramente foram analisadas as repetições de cada tratamento, com o objetivo de verificar se as curvas de sobrevivência entre as repetições não eram estatisticamente diferentes. Não ocorrendo diferenças significativas entre repetições, as três repetições de cada experimento foram agrupadas (“pooled”), dando-se início à análise das curvas de sobrevivência entre os tratamentos. Esses procedimentos foram implementados no software GraphPad Prism 3.0 (MOTULSKY, 1995).

A análise estatística de dados em Fitopatologia requer o desenvolvimento e, ou, adaptação de metodologias para o estudo de variáveis relacionadas com a técnica de Análise de Sobrevivência e também de padrões espaciais de doenças (BRANNEN et al., 2003). Além disso, considerando o volume de dados e o fato de que levantamentos para o acompanhamento de doenças são feitos regularmente ao longo do tempo, são necessários procedimentos de análise e geração de resultados e relatórios de forma automática, ou seja, mediante a utilização de softwares específicos. O software estatístico SAS[®] (Statistical Analysis System) dispõe de todos os métodos estatísticos geralmente utilizados em Análise de Sobrevivência, e, ainda, apresenta grande interatividade com usuários que não são especificamente da área estatística (ALLISON, 1995).

2.4. Importância do estudo da requeima do tomateiro

A família Solanaceae é de grande importância para a agricultura, destacando-se o tomate *Lycopersicon esculentum* (Mill) por sua versatilidade comercial. Contudo, os agricultores encontram muitas dificuldades devido à suscetibilidade da cultura a diversas doenças e pragas. Isso ocorre porque o tomate, produto de sucessivas seleções para que se chegasse a sua forma atual, teve sua variabilidade genética restringida (LABORY et al, 1999).

A requeima, causada por *Phytophthora infestans*, causa grandes perdas nos campos de cultivo, principalmente quando as condições ambientais são predisponentes, ou seja, alta umidade e baixa temperatura. A doença caracteriza-se por ser agressiva e de grande impacto destrutivo, podendo limitar ou, até mesmo, impedir o cultivo econômico do tomateiro sob condições favoráveis (LOPES; SANTOS, 1994; DEHNE; OERKE, 1998).

O patógeno apresenta rápida disseminação e colonização, podendo destruir a cultura em poucos dias, caso não sejam tomadas medidas de controle adequadas (TÖFOLI et al., 2003). A requeima pode ocorrer em qualquer fase do desenvolvimento da cultura, afetando severamente folhas, hastes, frutos e pecíolos que, em geral, apresentam aspecto semelhante à queima ou injúria por geadas (JONES et al., 1993; BLANCARD, 1996; KUROZAWA; PAVAN, 1997).

No Brasil, a requeima tornou-se importante para a cultura do tomateiro a partir dos anos 50, por ocasião da primeira grande expansão da cultura, exigindo a aplicação sistemática de fungicidas para o seu controle. Inicialmente, o controle foi realizado com base em fungicidas inespecíficos, como a calda Bordalesa, compostos cúpricos e ditiocarbamatos, que visavam ao controle conjunto do complexo requeima e pinta-preta (BOFF, 1988).

O uso de fungicidas é a principal medida de controle da doença, elevando, assim, os custos de produção. Dessa forma, torna-se necessário o conhecimento das possíveis fontes de resistência do tomate (SUINAGA et al., 2003), as quais podem ser encontradas em cultivares antigas, bem como nas espécies silvestres pertencentes ao mesmo gênero que se encontram registrados nos Bancos de Germoplasma (repositório onde se armazena a variabilidade genética de uma ou de várias espécies).

Esses bancos são importantes ferramentas para o melhoramento, pois são fontes essenciais de recursos gênicos (RIBEIRO et al., (2007). Geralmente, consiste de base física, em que o germoplasma é conservado, como centros ou instituições públicas e privadas que conservam as coleções sob a forma de sementes, explantes ou plantas a campo. Existem várias formas de conservação de germoplasma vegetal. Porém, a conservação de germoplasma de tomateiros é quase que exclusivamente realizado a campo, com poucas duplicatas de acessos de algumas espécies na forma de sementes ortodoxas e, ou, cultura de meristema *in vitro*.

3. MATERIAL E MÉTODOS

3.1. Descrição dos dados

O experimento foi conduzido no inverno de 2005, no Campo Experimental do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV). Foram utilizados acessos de tomateiros do BGH-UFV (Banco de Germoplasma de Hortaliças da UFV).

Sementes dos acessos foram semeadas em bandejas de 128 células, contendo substrato comercial e mantidas em casa de vegetação, sendo transplantadas para o campo aos 30 dias da semeadura. O delineamento experimental eleito foi o de blocos ao acaso, com três repetições e três plantas por parcela. Foram cultivados 42 acessos do BGH, sendo realizados todos os tratamentos culturais recomendados para a cultura.

Para o preparo do inoculo foram feitas coletas de folíolos infectados, em lavouras de tomate de cidades localizadas na Zona da Mata mineira, visto que fora detectado, entre os diversos isolados encontrados na região, grande variabilidade quanto à virulência (RIBEIRO et al., 2007). Os folíolos doentes foram colocados em bandejas previamente desinfetadas com álcool 70% e forradas com papel-toalha umedecido com água destilada, de modo a formar uma câmara úmida. As bandejas foram cobertas com saco plástico e mantidas a 18 °C. Após 24 h, as lesões dos folíolos foram cortadas e colocadas em um béquer contendo água destilada, agitados para uniformizar a suspensão de esporângios e levados à geladeira por 1 h. Para cada isolado, ajustou-se a concentração, com o hemacitômetro, para 1×10^3 esporângios/mL. Ao final, o inóculo obtido foi resultado da mistura de esporângios.

A inoculação foi realizada 60 dias após o transplante, e avaliou-se a severidade da doença a cada três dias, durante 18 dias. Ao final do experimento, obtiveram-se dados de sete avaliações temporais de cada acesso, estrutura essa que permitiu ajustar curvas de progressos de doença individualmente para cada repetição dentro de cada acesso.

3.2. Ajuste de curvas de progresso da doença e agrupamento dos acessos

O modelo exponencial $Y = y_0 \exp(rX)$ (CAMPBELL; MADEN, 1990) foi ajustado para descrever o comportamento da severidade (Y) em função do tempo (X), sendo y_0 a quantidade inicial de doença e r a taxa de progresso. Esse ajuste foi realizado para cada repetição de cada acesso, mediante o PROC MODEL (SAS, 1996), o que possibilitou a obtenção de um arquivo com as estimativas geradas para y_0 e r. Tal ajuste foi efetuado pelo método dos mínimos quadrados para modelos de regressão não-linear, utilizando o processo iterativo de Gauss-Newton. Depois desses ajustes, verificaram a convergência e os valores do coeficiente de correlação corrigido, os quais permitiram eliminar curvas anormais.

Realizou-se uma análise de variância multivariada utilizando o PROC GLM (SAS, 1996) com a opção MANOVA, com o intuito de avaliar a influência dos acessos do BGH sobre essas estimativas de y_0 e r por meio do critério de Roy (KHATTREE; NAIK, 2000). Sendo constatada essa influência, extraíram-se as médias ajustadas por quadrados mínimos (LSMEANS), as quais foram submetidas à análise de agrupamento no PROC CLUSTER (SAS, 1996) pelo método Centróide. Nesta última análise, obtiveram-se os valores dos desvios-padrão da raiz do quadrado médio (RMSSTD) em relação ao número de grupos (Figura 2), que disponibilizou a identificação do número ótimo de grupos em relação à máxima curvatura.

Para identificar o ponto de máxima curvatura e, conseqüentemente, o número ótimo de grupos (X_0), ajustou-se um modelo de regressão segmentada polinomial de segunda ordem (REZENDE et al., 2007). Para esse ponto, foi estimado um intervalo de confiança, do qual se utilizou o limite inferior como estimativa ótima do número de grupos.

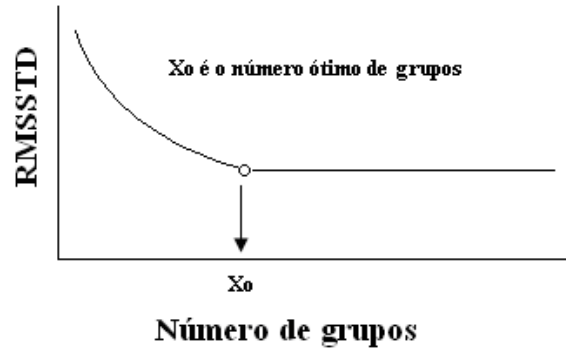


Figura 2 - Representação gráfica do comportamento do RMSSTD em função do número ótimo de grupos em relação à máxima curvatura.

Este modelo também é denominado Modelo Quadrático de Response Platô (MQRP), o qual assume que antes do ponto X_0 , tem-se uma parábola e depois uma reta constante. Assim, a descrição do modelo é a seguinte:

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i, & X_i \leq X_0 \\ P + e_i, & X_i > X_0 \end{cases}$$

em que: β_0 , β_1 e β_2 são os parâmetros da parábola ajustada antes de X_0 , e P é o platô.

3.3. Tempo até a severidade de 5% da doença (regressão inversa)

Neste trabalho, a severidade de 5% foi escolhida de acordo com Duarte et al. (2007). O tempo até a severidade de 5% ($X_{5\%}$) da doença foi computado via regressão inversa (Figura 3).

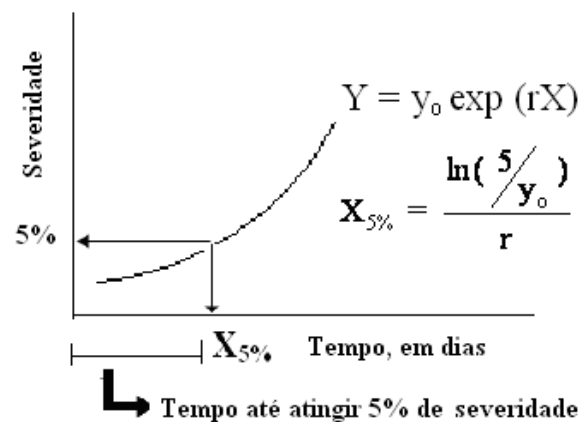


Figura 3 - Representação gráfica do processo de obtenção dos dados de sobrevivência por meio de regressão inversa.

Os valores de $X_{5\%}$ (variável dependente na análise de sobrevivência) foram calculados, considerando-se cada acesso e repetição dentro de cada um dos grupos formados. Para introduzir as censuras aos dados, uma vez que estes não as apresentavam naturalmente, confeccionou-se um intervalo de confiança de 95% para a média da variável resposta (tempo até atingir 5% de severidade). Dessa forma, valores dessa variável que se apresentavam acima do limite superior do intervalo de confiança foram considerados como observações censuradas.

3.4. Estimação de parâmetros na análise de sobrevivência

A função de sobrevivência pode ser estimada, considerando-se modelos paramétricos e técnicas não-paramétricas. A escolha de uma das abordagens não significa a exclusão da outra, por exemplo as técnicas não-paramétricas podem indicar evidências de que um modelo paramétrico não está adequado. Por isso, o uso das duas abordagens é essencial para garantir resultados fidedignos.

3.4.1. Técnicas não-paramétricas utilizadas na análise de sobrevivência

Em análise de sobrevivência, métodos não-paramétricos são usados quando o processo de estimação é realizado sem que se faça nenhuma suposição sobre a distribuição de probabilidade do tempo de sobrevivência, ou seja, não é imposto nenhum modelo teórico para as falhas e, ou, censuras observadas. Entre esses métodos são apresentados um correspondente à estimação de curvas de sobrevivência, o estimador de Kaplan-Meier e um relacionado com a comparação entre essas curvas, o teste *Logrank*.

3.4.1.1. Estimador de Kaplan-Meier

O estimador de Kaplan-Meier para a função de sobrevivência é o mais utilizado em estudos clínicos e vem ganhando cada vez mais espaço em estudos de confiabilidade e de processos biológicos (COLOSIMO; GIOLO, 2006). Esse estimador foi proposto por Kaplan e Meier (1958) para estimar a função de sobrevivência, sendo também chamado de estimador limite-produto. Ele é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ total de observações no estudo}} \quad (1)$$

$\hat{S}(t)$ 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.

O estimador de Kaplan-Meier, na sua construção, considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falha da amostra. Assim, para qualquer t , $S(t)$ pode ser escrito em termos de probabilidades condicionais. Suponha que existam n elementos no estudo e $k (\leq n)$ falhas distintas nos tempos $t_1 < t_2 < \dots < t_k$. Considerando $S(t)$ uma função discreta com probabilidade maior que zero somente nos tempos de falha $t_j, j = 1, \dots, k$, tem-se que:

$$S(t_j) = (1 - q_1)(1 - q_2) \dots (1 - q_j), \quad (2)$$

em que q_j é a probabilidade de um indivíduo morrer no intervalo $[t_{j-1}, t_j)$, sabendo que ele não morreu até t_{j-1} e considerando $t_0 = 0$. Ou seja, pode-se escrever q_j como:

$$q_j = P(T \in [t_{j-1}, t_j) | \geq t_{j-1}) \quad (3)$$

Dessa forma, a expressão geral de $S(t)$ é escrita em termos de probabilidades condicionais. O estimador de Kaplan-Meier se reduz, então, a estimar q_j que, adaptado da expressão (1), é dado por:

$$\hat{q}_j = \frac{\text{n}^\circ \text{ de falhas em } t_j}{\text{n}^\circ \text{ de observações sob risco em } t_{j-1}} = \frac{d_j}{n_j}, \text{ para } j = 1, \dots, k. \quad (4)$$

A expressão geral do estimador de Kaplan-Meier pode ser apresentada depois dessas considerações preliminares.

Considere:

$t_1 < t_2 \dots < t_k =$ os k tempos distintos e ordenados de falha;

$d_j =$ o número de falhas em $t_j, j = 1, \dots, k; e$

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 .

O estimador de Kaplan-Meier é, então, definido como:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right). \quad (5)$$

Uma justificativa simples para a expressão (5) do estimador de Kaplan-Meier vem da decomposição de $S(t)$ em termos dos q_j 's apresentada em (2). O estimador de Kaplan-Meier é obtido a partir de (5) se os q_j 's forem estimados por d_j/n_j , que foi expresso em palavras em (4). No artigo original, Kaplan e Meier (1958) justificaram a expressão (5), mostrando que ela é o estimador de máxima verossimilhança de $S(t)$. Os principais passos dessa prova são indicados a seguir. Suponha, como feito anteriormente, que d_j observações falham no tempo t_j , para $j = 1, \dots, k$, e m_j observações são censuradas no intervalo $[t_j, t_{j+1})$, nos tempos t_{j1}, \dots, t_{jmj} . A probabilidade de falha no tempo t_j é, então,

$$S(t_j) - S(t_{j+}),$$

com $S(t_{j+}) = \lim_{\Delta t \rightarrow 0^+} S(t_j + \Delta t)$, $j = 1, \dots, k$. No entanto, a contribuição para a função de verossimilhança de um tempo de sobrevivência censurado em $t_{j\ell}$ para $\ell = 1, \dots, m_j$, é: $P(T > t_{j\ell}) = S(t_{j\ell+})$

A função de verossimilhança pode, então, ser escrita como:

$$L(S(.)) = \prod_{j=0}^k \left\{ \left[S(t_j) - S(t_{j+}) \right]^{d_j} \prod_{\ell=1}^{m_j} S(t_{j\ell+}) \right\} \quad (6)$$

Pode-se mostrar que $S(t)$, que maximiza $L(S(.))$, é exatamente a expressão (5). Esta definição do estimador de máxima verossimilhança é uma generalização do conceito usual utilizado em modelos paramétricos, em que se tem tantos parâmetros quanto falhas distintas. Entretanto, o resultado de problemas como este, em que muitos parâmetros estão envolvidos, deve ser tratado com cuidado. Detalhes dessa prova são encontrados em Kalbfleisch e Prentice (1980), os quais relataram, ainda, que, devido a essa propriedade, o estimador Kaplan-Meier pode ser considerado não-viciado para amostras grandes e de fácil convergência assintótica para um processo gaussiano (distribuição normal).

Para que se possam construir intervalos de confiança e testar hipóteses para $S(t)$, é necessário, no entanto, avaliar a precisão do estimador de Kaplan-Meier. Esse estimador, assim como outros, está sujeito a variações que devem ser descritas em termos de estimações intervalares. A expressão para a variância assintótica desse estimador é dada por:

$$\hat{V}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (7)$$

Esta expressão (7) pode ser obtida a partir de propriedades do estimador de máxima verossimilhança. Como $S(t)$, para t fixo tem distribuição assintótica Normal: segue que um intervalo aproximado de $100(1-\alpha)\%$ de confiança para $S(t)$ é dado por:

$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\hat{Var}(\hat{S}(t))}$, em que $\alpha/2$ denota o percentil da distribuição Normal padrão.

Entretanto, para valores extremos de t , esse intervalo de confiança pode apresentar limite inferior negativo ou limite superior maior do que 1. Nesses casos, o problema é resolvido utilizando-se uma transformação para $S(t)$, como $\hat{U}(t) = \log[-\log(\hat{S}(t))]$, sugerida por Kalbfleish e Prentice (1980), que tem variância assintótica estimada por:

$$\hat{V}(\hat{U}(t)) = \frac{\sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}}{\left[\sum_{j:t_j < t} \log\left(\frac{n_j - d_j}{n_j}\right) \right]^2} = \frac{\sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}}{\left[\log(\hat{S}(t)) \right]^2} \quad (8)$$

Assim, um intervalo aproximado de $100(1-\alpha)\%$ de confiança para $S(t)$ é dado por:

$$\left[\hat{S}(t) \right]^{\exp\left\{ \pm z_{\alpha/2} \sqrt{\hat{Var}(\hat{U}(t))} \right\}}, \text{ que assume valores no intervalo } [0,1].$$

3.4.1.2. Teste *Logrank* para comparação de curvas de sobrevivência

O teste *Logrank* (MANTEL, 1966) é o mais usado em análise de sobrevivência e é particularmente apropriado quando a razão das funções de risco dos grupos a serem comparados é aproximadamente constante. Isto é, as populações têm a propriedade de

riscos proporcionais. A estatística desse teste é a diferença entre o número observado de falhas em cada grupo e uma quantidade que, para muitos propósitos, pode ser pensada como o correspondente número esperado de falhas sob a hipótese nula. A expressão do teste *Logrank* é obtida de forma similar à do conhecido teste de Mantel-Haenszel (1959), para combinar tabelas de contingência. O teste *Logrank* tem, também, a mesma expressão do teste escore para o modelo de regressão de Cox.

Considere, inicialmente, o teste de igualdade de duas funções de sobrevivência $S_1(t)$ e $S_2(t)$. Sejam $t_1 < t_2 < \dots < t_k$ os tempos de falha distintos da amostra formada pela combinação das duas amostras individuais. Suponha que no tempo t_j aconteçam d_j falhas e que n_j indivíduos estejam sob risco em um tempo imediatamente inferior a t_j na amostra combinada e, respectivamente, d_{ij} e n_{ij} na amostra i ; $i = 1, 2$ e $j = 1, \dots, k$. Em cada tempo de falha t_j , os dados podem ser dispostos em forma de uma tabela de contingência 2 x 2 com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i . (Tabela 1).

Tabela 1 - Tabela de contingência gerada no tempo t_j para dois grupos individuais

	Grupo 1	Grupo2	Total
Falha	d_{1j}	d_{2j}	d_j
Não falha	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
Total	n_{1j}	n_{2j}	n_j

Condicional à experiência de falha e censura até o tempo t_j (fixando as marginais de coluna) e ao número de falhas no tempo t_j (fixando as marginais de linha), a distribuição de d_{2j} é, então, uma hipergeométrica:

$$\frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}}$$

A média de d_{2j} é $w_{2j} = n_{2j} d_j n_j^{-1}$, o que equivale dizer que, se não houver diferença entre as duas populações no tempo t_j , o número total de falhas (d_j) pode ser dividido entre as duas amostras de acordo com a razão entre o número de indivíduos sob risco em cada amostra e o número total sob risco. A variância de d_{2j} obtida a partir da distribuição hipergeométrica é:

$$(V_j)_{22} = n_{2j}(n_j - n_{2j}) d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}$$

Então, a estatística $d_{2j} - w_{2j}$ tem média zero e variância $(V_j)_{22}$. Se as k tabelas de contingência forem independentes, um teste aproximado para a igualdade das duas funções de sobrevivência pode ser baseado na estatística:

$$T = \frac{\left[\sum_{j=1}^k (d_{2j} - w_{2j}) \right]^2}{\sum_{j=1}^k (V_j)_{22}} \quad (9)$$

que, sob a hipótese nula $H_0: S_1(t) = S_2(t)$ para todo t no período de acompanhamento, tem uma distribuição qui-quadrado com 1 grau de liberdade para grandes amostras.

A generalização do teste *Logrank* para a igualdade de $r > 2$ funções de sobrevivência $S_1(t), \dots, S_r(t)$ não é complicada. Considere a mesma notação anterior, com índice i variando, agora, entre 1 e r . Dessa forma, os dados podem ser arranjados em forma de uma tabela de contingência $2 \times r$ com d_{ij} falhas e $n_{ij} - d_{ij}$ sobreviventes na coluna i , ou seja, a Tabela 1 passaria a ter r colunas em vez de simplesmente duas.

Condicional à experiência de falha e censura até o tempo t_j , a distribuição conjunta de d_{2j}, \dots, d_{rj} é, então, uma hipergeométrica multivariada, isto é,

$$\frac{\prod_{i=1}^r \binom{n_{ij}}{d_{ij}}}{\binom{n_j}{d_j}}$$

A média de d_{ij} é $w_{ij} = n_{ij} d_j n_j^{-1}$, a variância de d_{ij} e a co-variância de d_{ij} e d_{lj} são, respectivamente, $(V_j)_{ii} = n_{ij}(n_j - n_{ij}) d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}$ e $(V_j)_{il} = -n_{ij} n_{lj} d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}$. Então, a estatística $v' = (d_{2j} - w_{2j}, \dots, d_{rj} - w_{rj})$ tem média zero e matriz de variância-co-variância V_j de dimensão $r - 1$, com $(V_j)_{ii}, i=2, \dots, r$, na diagonal principal e os elementos $(V_j)_{il}, l=2, \dots, r$ fora da diagonal principal. Pode-se, então, formar a estatística v , somando sobre todos os tempos distintos de falha, isto é:

$$v = \sum_j^k v_j$$

com v um vetor de dimensão $(r - 1) \times 1$, cujos elementos são as diferenças entre os totais observados e esperados de falha.

Considerando novamente a suposição de que as k tabelas de contingência são independentes, a variância da estatística v será $V = V_1 + \dots + V_k$. Um teste aproximado para a igualdade das r funções de sobrevivência pode ser baseado na seguinte estatística:

$$T = v'V^{-1}v$$

que, sob H_0 (igualdades das curvas), tem distribuição qui-quadrado com $r - 1$ graus de liberdade para amostras grandes. Os graus de liberdade são $r - 1$ e não r , pois os elementos de v somam zero.

3.4.2. Modelos probabilísticos utilizados na análise de sobrevivência

Apesar de existirem técnicas não paramétricas para estimar parâmetros em análise de sobrevivência, há também a opção de se ajustarem os dados utilizando modelos paramétricos. Esses modelos são chamados modelos paramétricos de probabilidade para o tempo de falha (COLOSIMO, 1995). Na literatura existem vários autores que fazem uma apresentação exaustiva desses modelos e que podem ser usados pelo pesquisador, entre eles podemos citar Johnson e Kotz (1970), Lawless (1982), Marubini e Valsecchi (1995).

Embora exista uma série de modelos probabilísticos em análise de sobrevivência, alguns desses ocupam posição de destaque por sua comprovada adequação a várias situações práticas, sendo usados com bastante frequência. Os principais modelos são: Exponencial, Weibull e Log-normal. Os modelos paramétricos têm sido utilizados por pesquisadores em problemas práticos com bastante sucesso, mesmo com a presença de fatores de risco (BOLFARINE et al., 1991).

3.4.2.1. Modelo exponencial

A distribuição exponencial de um parâmetro, por ter uma estrutura simples, é muito utilizada. Se T tem distribuição exponencial com parâmetro $\alpha > 0$, a função densidade de probabilidade, $f(t)$, é dada por:

$$f(t) = \left(\frac{1}{\alpha}\right) \exp\left(\frac{-t}{\alpha}\right),$$

em que $t \geq 0$, em que o parâmetro $\alpha \geq 0$ é o tempo médio de o produto não falhar e tem a mesma unidade do tempo de falha t (se t é medido em dias, α também será). A função de sobrevivência referente a esta distribuição, $S(t)$, é representada pela seguinte expressão:

$$S(t) = \exp\left(\frac{-t}{\alpha}\right), t \geq 0.$$

A função de risco, $h(t)$, é constante, sendo esta a principal característica dessa distribuição (COLOSIMO, 1995). Isso significa que tanto uma unidade velha quanto uma nova, que ainda não falharam, têm o mesmo risco de fazê-lo em um intervalo futuro. Essa propriedade é chamada de falta de memória da distribuição exponencial. A referida função é representada por:

$$h(t) = \frac{1}{\alpha}, t \geq 0.$$

A forma típica dessas três funções varia com os valores de α . (Figura 3)

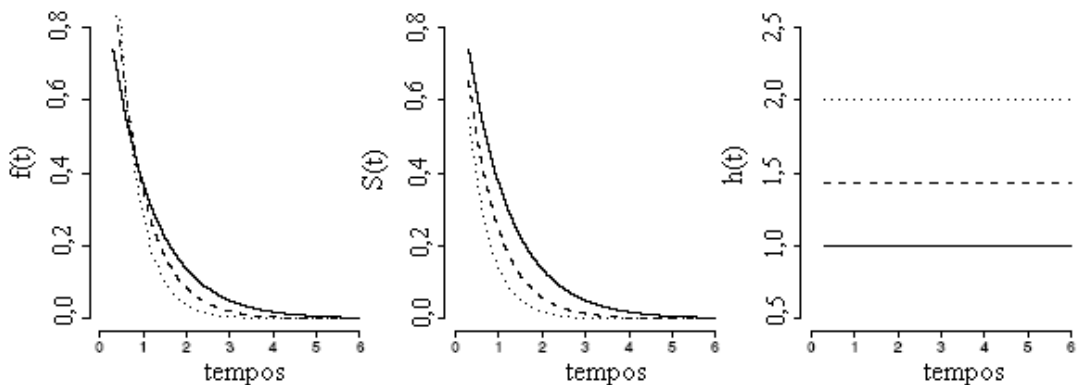


Figura 3 - Forma típica das funções de densidade de probabilidade $f(t)$, de sobrevivência $S(t)$ e de taxa de falha $h(t)$ da distribuição exponencial de $\alpha = 1,0$ (—), $0,7$ (- - -) e $0,5$ (....).

3.4.2.2. Modelo Weibull

A distribuição de Weibull (1939) representa uma generalização da distribuição exponencial e, de acordo com Lawless (1982), é bastante utilizada no ajuste de dados de confiabilidade nas diversas áreas do conhecimento, entre elas Medicina e Engenharia. Na engenharia, a distribuição Weibull é a principal função de confiabilidade, sendo utilizada para modelar a distribuição da vida útil e taxa de risco em produtos industriais (XIE, 2002; KIM, 2003; MUDHOLKAR, 1993; LAI, 1993). É adequada para produtos formados de várias partes, cuja falha ocorre quando a primeira parte falhar. Se T tem distribuição de Weibull com parâmetros $\alpha > 0$ e $\gamma > 0$, a função densidade de probabilidade, $f(t)$, é dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \quad t \geq 0$$

Os parâmetros de forma γ e o de escala α são positivos. O parâmetro α tem a mesma unidade de t , e γ não tem unidade. Se $\gamma = 1$, tem-se a distribuição exponencial, sendo, assim, a distribuição exponencial é um caso particular da distribuição de Weibull.

A função de sobrevivência, $S(t)$, é dada por:

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

A função risco, $h(t)$, é definida como: $h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}$, em que:

$$h(t) \begin{cases} \text{Estritamente crescente para } \gamma > 1; \\ \text{Estritamente decrescente para } \gamma < 1; \text{ e} \\ \text{constante para } \gamma=1 \text{ (exponencial).} \end{cases}$$

A sua popularidade deve-se ao fato de poder apresentar-se sob várias formas, todas com uma propriedade básica que é a sua função risco monótona, ou seja, ela é crescente, decrescente ou constante. Collet (1994) afirmou que o modelo Weibull é tão importante para a análise de confiabilidade quanto à distribuição normal dos modelos lineares.

As formas das funções de densidade, de sobrevivência e de taxa de falha (risco) de uma variável T com distribuição de Weibull variam com alterações nos valores dos parâmetros γ e α (Figura 4).

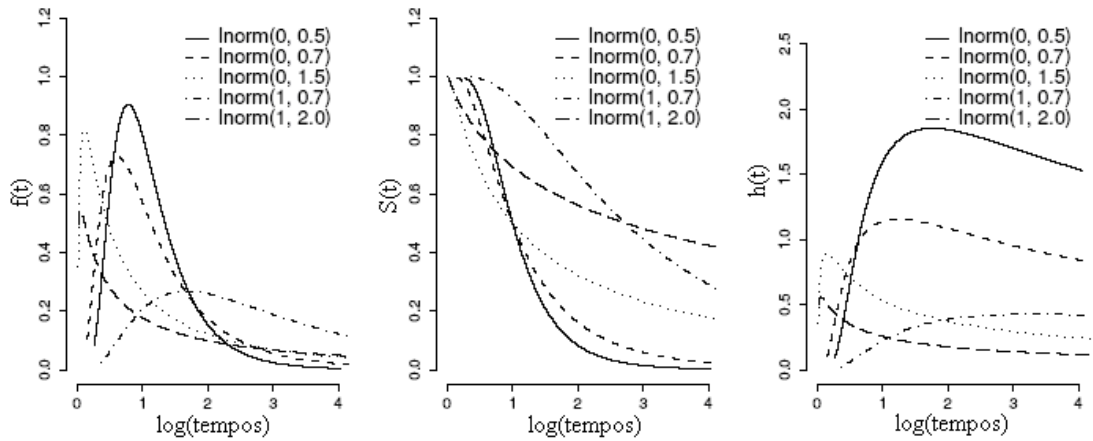


Figura 4 - Forma típica das funções de densidade de probabilidade $f(t)$, de sobrevivência $S(t)$ e de taxa de falha $h(t)$ da distribuição de Weibull para alguns valores dos parâmetros (γ e α).

3.4.2.3. Modelo Log-normal

A distribuição Log-normal é muito usada para ajustar dados referentes à confiabilidade, assim como a distribuição Weibull. Existem diversas aplicações deste modelo em testes do tempo de falha de produtos (NELSON, 1990). Uma discussão detalhada sobre esse modelo pode ser encontrada em Crow e Shimizu (1988). Se T tem distribuição Log-normal com função densidade de probabilidade, $f(t)$, representada por:

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\} \quad t \geq 0$$

em que μ é a média do logaritmo do tempo de falha e σ , o desvio padrão.

As funções de sobrevivência e de taxa de falha de uma variável Log-normal não apresentam uma forma analítica explícita e são, desse modo, expressas, respectivamente, por

$$S(t) = \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right) \quad \text{e} \quad \lambda(t) = \frac{f(t)}{S(t)}$$

em que Φ é a função de distribuição acumulada de uma normal padrão.

Pode ser observado que as funções de taxa de falha não são monótonas como as da distribuição de Weibull. Elas crescem, atingem um valor máximo e depois decrescem (Figura 5).

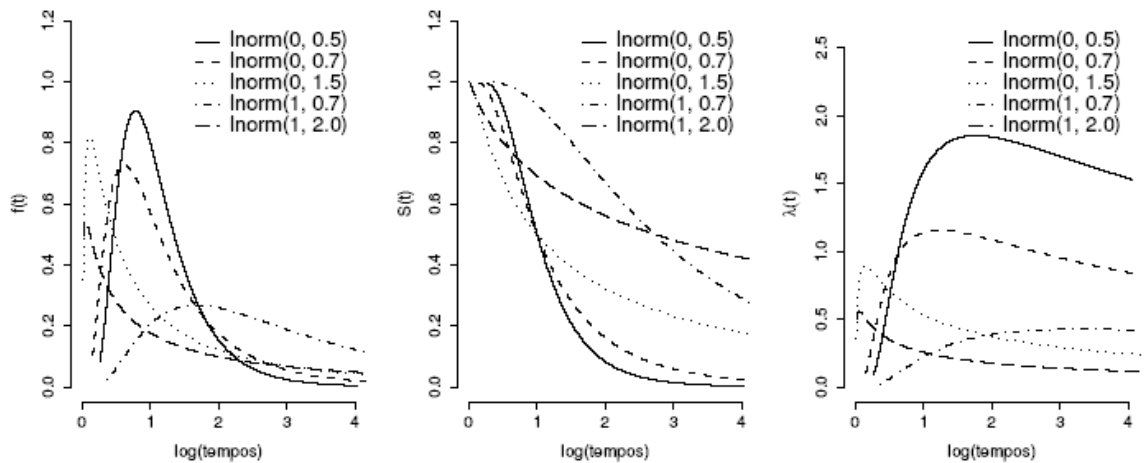


Figura 5 - Forma típica das funções de densidade de probabilidade $f(t)$, de sobrevivência $S(t)$ e de taxa de falha $h(t)$ da distribuição Log-normal de alguns valores dos parâmetros (μ e σ).

Existem outras distribuições de probabilidade apropriadas para modelar o tempo de falha de produtos, materiais e situações clínicas. Entre elas, podem ser citadas as distribuições logística, Log-logística, gama, Log-gama, Rayleigh, normal inversa e Gompertz. Diversos textos apresentam, ainda, a popular função de taxa de falha do tipo “banheira” que descreve o comportamento das taxas de falhas de certos produtos industriais e, principalmente, do tempo de vida dos seres humanos (NELSON, 1990).

3.4.2.4. Modelo Gama generalizado

Outra distribuição que merece destaque em análise de sobrevivência é a distribuição Gama generalizada. Essa distribuição foi introduzida por Stacy (1962) e é caracterizada por três parâmetros, γ , k e α , todos positivos. Sua função de densidade é dada por:

$$f(t) = \frac{\gamma}{\Gamma(k)} \frac{t^{\gamma k - 1}}{\alpha^{\gamma k}} \exp \left\{ - \left(\frac{t}{\alpha} \right)^{\gamma} \right\} \quad t > 0$$

em que $\Gamma(k)$ é a função Gama, isto é, $\Gamma(k) = \int_0^{\infty} x^{k-1} \exp(-x) dx$. Para essa distribuição, têm-se um parâmetro de escala, α , e dois de formas γ e k , o que a torna bastante flexível.

Note-se a partir da função de densidade da distribuição Gama generalizada, que:

- i) para $\gamma = k = 1$, tem-se $T \sim \text{Exp}(\alpha)$;
- ii) para $k = 1$, tem-se $T \sim \text{Weibull}(\gamma, \alpha)$; e
- iii) para $\gamma = 1$, tem-se $T \sim \text{Gama}(k; \alpha)$.

Pode-se, ainda, mostrar (LAWLESS, 1982) que a distribuição Log-normal aparece como um caso-limite da distribuição Gama generalizada quando $k \rightarrow \infty$.

Do que foi exposto, tem-se que a distribuição Gama generalizada inclui como casos especiais as distribuições: exponencial, Weibull, Gama e Log-normal. Essa propriedade da Gama generalizada faz que ela mesma seja de grande utilidade, por exemplo na comparação entre modelos probabilísticos alternativos.

3.4.3. Estimação de parâmetros dos modelos probabilísticos

Os modelos probabilísticos apresentados anteriormente são caracterizados por quantidades geralmente desconhecidas, denominadas parâmetros. O modelo Gama generalizado é caracterizado por três parâmetros, os modelos de Weibull, Log-normal por dois parâmetros e o exponencial por apenas um. Essas quantidades conferem uma forma geral aos modelos probabilísticos. Entretanto, em cada estudo envolvendo tempos de falha, os parâmetros devem ser estimados a partir das observações amostrais para que o modelo fique determinado e, assim, seja possível responder às perguntas de interesse.

Existem alguns métodos de estimação conhecidos na literatura estatística. Talvez o mais conhecido seja o método de mínimos quadrados. No entanto, esse método é inadequado para estudos de tempo de vida. A principal razão é a sua incapacidade de incorporar censuras no seu processo de estimação. O método de máxima verossimilhança surge como uma opção apropriada para esse tipo de dados. Ele incorpora as censuras, é relativamente simples de ser entendido e possui propriedades ótimas para grandes amostras. Resumindo, pode-se dizer que nos modelos paramétricos a função de sobrevivência depende de um vetor de parâmetros θ , que pode ser estimado via máxima verossimilhança, ou seja, os estimadores de máxima verossimilhança para θ são obtidos maximizando o logaritmo natural da função de verossimilhança.

3.4.4. Escolha do modelo probabilístico

A escolha do modelo a ser utilizado é um tópico extremamente importante na análise paramétrica de dados de tempo de vida. O método de máxima verossimilhança só pode ser aplicado depois de ter sido definido um modelo probabilístico adequado para os dados. Por exemplo, somente após ter definido que o modelo Log-normal se ajusta bem aos dados é que o método de máxima verossimilhança pode ser usado para estimar μ e σ . Entretanto, se o modelo Log-normal for usado inadequadamente para certo conjunto de dados, toda a análise estatística fica comprometida e, conseqüentemente, as respostas às perguntas de interesse ficam distorcidas. Dessa forma, como definir o melhor modelo?

Algumas vezes, existem evidências provenientes de testes realizados no passado de que certo modelo se ajusta bem aos dados. No entanto, em muitas situações esse tipo de informação não está disponível. A solução para essas situações é basicamente empírica. Sabe-se que as distribuições apresentadas anteriormente são típicas para dados de tempos de vida. A proposta empírica consiste em ajustar os modelos probabilísticos apresentados (Exponencial, Weibull etc.) e, com base na comparação entre valores estimados e observados, decidir qual deles “melhor” explica os dados amostrais. A forma mais simples e eficiente de selecionar o “melhor” modelo a ser usado para um conjunto de dados é através de técnicas gráficas.

As técnicas gráficas são extremamente úteis na seleção de modelos. Entretanto, as conclusões a partir delas podem diferir para diferentes analistas, ou seja, existe nas técnicas gráficas um componente subjetivo na sua interpretação.

Outra forma de discriminar modelos é por meio de testes de hipóteses com modelos encaixados. Nesse caso, a conclusão é direta e, portanto, não envolve qualquer componente subjetivo na sua interpretação.

Um teste bastante usual é o que utiliza a estatística da razão de verossimilhanças (TRV) em modelos encaixados (COX; HINKLEY, 1974). As hipóteses a serem testadas geralmente são: H_0 : O modelo de interesse é adequado *versus* H_a : O modelo não é adequado. Para a realização deste teste, deve ser identificado um modelo generalizado tal que os modelos de interesse são casos particulares. O teste é realizado a partir dos seguintes dois ajustes: (1) modelo generalizado e obtenção do valor do logaritmo de sua função de verossimilhança ($\ln L(\hat{\theta}_G)$); (2) modelo de

interesse e obtenção do valor do logaritmo de sua função de verossimilhança ($\ln L(\hat{\theta}_M)$). A partir desses valores, é possível calcular a estatística da razão de verossimilhança, isto é:

$$\text{TRV} = -2 \log \left[\frac{L(\hat{\theta}_M)}{L(\hat{\theta}_G)} \right] = 2 [\log L(\hat{\theta}_G) - \log L(\hat{\theta}_M)]$$

que, sob H_0 , tem aproximadamente uma distribuição qui-quadrado com graus de liberdade igual à diferença do número de parâmetros ($\hat{\theta}_G$ e $\hat{\theta}_M$) dos modelos comparados.

No contexto de análise de sobrevivência, esse teste é usualmente realizado utilizando-se a distribuição Gama generalizada que apresentam os modelos: exponencial, Weibull, Log-normal e Gama, como modelos encaixados, uma vez que todos eles, como vistos anteriormente, são casos especiais do modelo Gama generalizado.

3.4.5. Métodos para comparação dos parâmetros dos modelos probabilísticos

Em se tratando de testes de hipóteses referentes à igualdade dos parâmetros de diferentes curvas de sobrevivência determinadas por modelos paramétricos, uma forma simples e objetiva é criar intervalos de confiança para cada parâmetro de cada curva que representa os grupos comparados. Assim, a sobreposição ou não desses intervalos de confiança (IC) trará novas informações sobre a diferença ou não dos parâmetros dos grupos sem que seja necessário executar quaisquer testes formais, uma vez que esses geralmente exigem pressuposições (SOUZA, 1998). Porém, de qualquer forma, Carvalho et al. (2005) relataram que, para se utilizar tal método, a distribuição dos parâmetros não deve estar muito distante de uma normal e, ou, o tamanho de amostra deve ser teoricamente grande.

Para que as diferenças sejam significativas, o limite inferior do IC para um parâmetro de uma curva, representativa de um grupo de acesso, deve ser maior que o limite superior do IC para esse mesmo parâmetro de outra curva referente a outro grupo, ou seja, não deve haver sobreposição da amplitude desses intervalos. Dessa forma, se uma curva de sobrevivência for estimada para cada grupo de interesse, e se o modelo usado apresentar dois parâmetros, por exemplo, escala e forma, o fato de apenas um desses parâmetros mostrar-se diferente entre duas curvas comparadas já basta para

afirmar que essas são estatisticamente diferentes ao nível α de probabilidade considerada na confecção do IC.

3.5. Softwares utilizados para a análise de sobrevivência

Para a implementação dos métodos não-paramétricos de análise de sobrevivência, caracterizados pelo estimador Kaplan-Meyer e pelo teste *Logrank*, utilizou-se o software estatístico SAS (SAS, 1996), por meio do procedimento LIFETEST. Devido à maior facilidade em relação à comparação de modelos, os métodos paramétricos referentes ao ajuste dos modelos probabilísticos e à obtenção dos intervalos de confiança a 95% dos seus parâmetros, utilizou-se o pacote SURVREG, do software estatístico R (R DEVELOPMENT CORE TEAM, 2008).

4. RESULTADOS E DISCUSSÃO

4.1. Ajuste de curvas de progresso da doença e agrupamento dos acessos

O modelo exponencial $Y = y_0 \exp (rX)$ foi ajustado para descrever o comportamento da severidade da requeima causada por *Phytophthora infestans* (Y) em função do tempo (X). Na Tabela 2 são apresentados os valores estimados de y_0 e r desse modelo para cada repetição de cada acesso de tomateiro.

Valores da severidade da doença que foram superiores a 5% foram excluídos do banco de dados por gerarem tempos negativos na regressão inversa. Nessa exclusão, o acesso de número 13 foi perdido, pois suas três repetições apresentaram valores acima de 5%. Assim, as análises continuaram com apenas 41 acessos e com número de repetições diferentes.

Realizada a análise de variância multivariada, verificou-se que havia influência dos acessos sobre as estimativas do modelo utilizando o critério de Roy (Tabela 3). Após esta constatação, foram estimadas as médias de y_0 e r de cada acesso por quadrados mínimos para serem utilizadas na análise de agrupamento.

Na análise de agrupamento dos acessos, utilizou-se o método do Centróide, sendo obtidos os valores dos desvios-padrão da raiz do quadrado médio (RMSSTD) em relação ao número de grupos, gerou-se, assim, um gráfico (Figura 6), que disponibilizou a identificação do número ótimo de grupos em relação à máxima curvatura. Esse número foi obtido por meio do limite inferior do intervalo de confiança [10,23 : 22,13].

Com a obtenção do número ótimo, os 41 acessos foram agrupados em 10 grupos (Tabela 4), representando curvas de progressos estatisticamente diferentes, e esses acessos correspondem aos diferentes tratamentos a serem considerados na análise

de sobrevivência. Esse método foi empregado com o intuito de oferecer um critério objetivo para a escolha do número de grupos e não simplesmente adotar critérios subjetivos como o faz a maioria dos trabalhos nesta área.

Tabela 2 - Modelo exponencial $Y = y_0 \exp(rX)$ ajustado para cada repetição dos 42 acessos, afim descrever o comportamento da severidade (Y) em função do tempo (X) na queima do tomateiro

Acesso	Rep	y_0	r	Acesso	Rep	y_0	r	Acesso	Rep	y_0	r
1	1	2,3424	0,1843	15	1	4,3570	0,1729	29	1	0,5806	0,2490
1	2	0,7922	0,2413	15	2	6,0125	0,1254	29	2	1,8313	0,1810
1	3	3,1414	0,1693	15	3	4,0101	0,1556	29	3	0,7386	0,2414
2	1	0,4806	0,2656	16	1	4,9843	0,1623	30	1	1,5835	0,1779
2	2	2,7851	0,1717	16	2	4,1067	0,1579	30	2	0,7153	0,2271
2	3	3,5370	0,1587	16	3	11,9009	0,1245	30	3	4,0332	0,1373
3	1	2,4472	0,1826	17	1	1,2078	0,1838	31	1	2,8913	0,1497
3	2	4,9028	0,1345	17	2	1,3866	0,1746	31	2	1,8604	0,1919
3	3	2,5814	0,1943	17	3	4,0144	0,1828	31	3	4,9802	0,1573
4	1	1,7374	0,1993	18	1	3,7340	0,1340	32	1	3,5986	0,1482
4	2	3,4530	0,1935	18	2	1,0684	0,2260	32	2	2,5279	0,2053
4	3	2,8063	0,1833	18	3	1,1990	0,2150	32	3	10,8265	0,1156
5	1	5,4483	0,1455	19	1	2,7700	0,1714	33	1	5,2862	0,1535
5	2	2,1563	0,1939	19	2	4,1027	0,1743	33	2	1,7403	0,1803
5	3	3,8578	0,1497	19	3	2,6030	0,1747	33	3	2,6197	0,1623
6	1	0,3033	0,2319	20	1	1,1169	0,2256	34	1	1,9658	0,1945
6	2	6,6656	0,1260	20	2	4,1284	0,1563	34	2	4,4890	0,1574
6	3	0,9496	0,2116	20	3	2,1025	0,1710	34	3	2,2095	0,1791
7	1	2,5253	0,1675	21	1	4,7347	0,1325	35	1	7,0537	0,1493
7	2	7,6295	0,1507	21	2	0,3813	0,2928	35	2	1,3175	0,2039
7	3	1,3148	0,2358	21	3	2,4211	0,1627	35	3	3,2694	0,1830
8	1	0,9012	0,1457	22	1	1,4136	0,2244	36	1	6,1416	0,1564
8	2	1,5057	0,1702	22	2	2,2044	0,2058	36	2	3,6838	0,1857
8	3	1,0077	0,2136	22	3	5,6703	0,1610	36	3	5,9606	0,1624
9	1	0,4954	0,2648	23	1	3,1388	0,1660	37	1	1,3116	0,1948
9	2	1,3572	0,2082	23	2	3,0462	0,1835	37	2	1,6851	0,1727
9	3	4,5071	0,1492	23	3	5,3808	0,1564	37	3	1,5514	0,2245
10	1	1,3928	0,1905	24	1	3,7287	0,1502	38	1	2,3824	0,2059
10	2	4,3055	0,1330	24	2	5,6535	0,1673	38	2	1,4348	0,1993
10	3	1,0471	0,2111	24	3	3,5555	0,1518	38	3	6,7702	0,1551
11	1	1,3214	0,2157	25	1	2,9364	0,1580	39	1	0,3917	0,2797
11	2	5,9358	0,1567	25	2	0,6548	0,2421	39	2	1,7549	0,2008
11	3	0,5519	0,2374	25	3	0,1592	0,3349	39	3	1,6987	0,2007
12	1	1,5038	0,1918	26	1	1,2820	0,2134	40	1	5,3253	0,1349
12	2	1,4189	0,2316	26	2	6,7796	0,1393	40	2	2,6882	0,1635
12	3	11,3479	0,1309	26	3	2,3712	0,1604	40	3	4,3144	0,1493
13	1	6,3375	0,1535	27	1	7,7668	0,1325	41	1	1,1228	0,2158
13	2	21,6176	0,0962	27	2	4,6494	0,1654	41	2	2,3682	0,1705
13	3	6,7114	0,1341	27	3	5,3687	0,1482	41	3	1,5940	0,2051
14	1	6,5885	0,1440	28	1	0,3430	0,2963	42	1	11,2703	0,1312
14	2	1,3169	0,2065	28	2	4,6970	0,1208	42	2	2,6237	0,1714
14	3	6,4665	0,1526	28	3	7,3096	0,1509	42	3	4,6991	0,1720

y_0 : quantidade inicial de doença e r: taxa de progresso da doença.

Tabela 3 - Critério de Roy no teste da análise multivariada e aproximação F para a hipótese de não haver efeito dos acessos sobre as estimativas do modelo

Valor	F	GL	Pr > F
1,22278936	1,74	40	0,0268

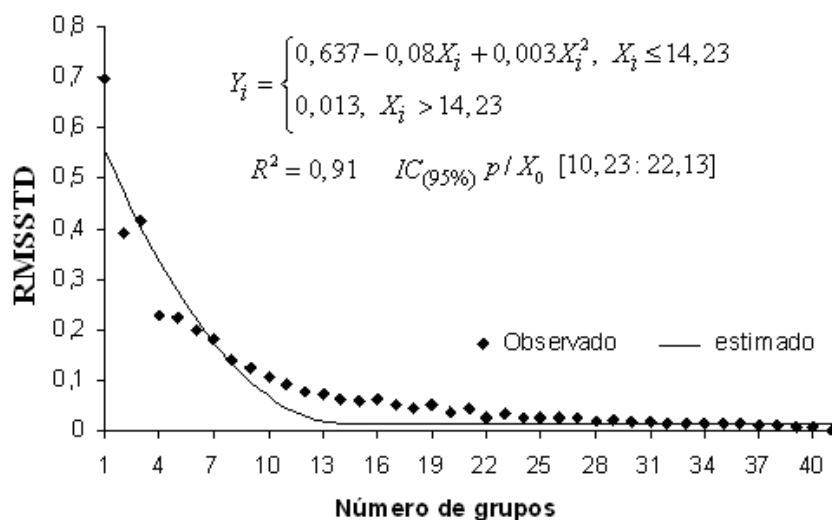


Figura 6 - Comportamento dos valores dos desvios-padrão da raiz do quadrado médio (RMSSTD), disponibilizando o número ótimo de grupos em relação à máxima curvatura.

Tabela 4 - Acessos de tomateiro reunidos em 10 grupos pelo método do centróide

Grupo	Freqüência	Acessos(repetições)
I	15	3(3), 19(3), 23(2), 31(3), 32(2), 40(2)
II	11	4(3), 20(3), 21(3), 28(2)
III	5	5(2), 34(3)
IV	33	1(3), 2(3), 7(2), 9(3), 10(3), 17(3), 18(3), 22(2), 26(2), 30(3), 33(2), 35(2), 38(2)
V	3	16(2), 27(1)
VI	15	8(3), 11(2), 14(1), 25(3), 29(3), 39(3)
VII	5	24(2), 36(1), 42(2)
VIII	8	12(2), 37(3), 41(3)
IX	2	15(2)
X	2	6(2)

Na Figura 7 são apresentadas as curvas de progresso da doença para os valores médios de y_0 e r dos acessos de tomateiro reunidos em 10 grupos. O acesso BGH-6,

pertencente ao grupo 10, obteve um progresso de doença menor que os demais, caracterizando-se, assim, sua maior resistência à doença. Também é relevante ressaltar que tal característica é tão marcante que esse acesso foi alocado individualmente em grupo. De forma contrária, observou-se que nos acessos pertencentes ao grupo 5 (BGH-16 e BGH-27) a doença teve progresso, uma vez que as curvas geradas são mais acentuadas, permitindo assim, mediante essa análise, caracterizar sua maior suscetibilidade.

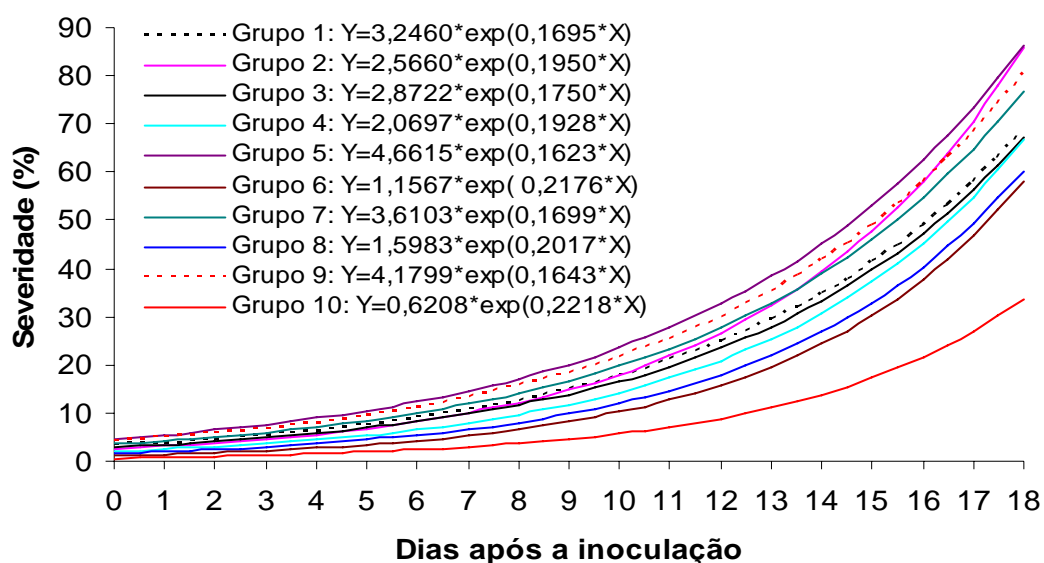


Figura 7 - Curvas de progresso da requeima do tomateiro em 41 acessos reunidos em 10 grupos.

4.2. Tempo até a incidência de 5% da doença (regressão inversa)

Via regressão inversa, os valores de tempo para atingir severidade de 5% foram calculados (Tabela 5). Aos tempos estimados foram incorporadas as censuras, uma vez que esses não as apresentavam naturalmente. Para tanto, calculou-se um intervalo de confiança de 95% [4,1143 : 5,2445] para a média dos tempos (4,6794) com erro padrão de 2,8333. Na Tabela 6 são apresentados os tempos (dias) até a ocorrência de 5% da doença, mostrando as falhas e censuras dos dados utilizados na análise de sobrevivência.

Tabela 5 - Valores estimados de y_0 e r que foram empregados para calcular o tempo até a severidade de 5% da doença, utilizando a regressão inversa

Acesso	Repetição	y_0	r	Tempo	Acesso	Repetição	y_0	r	Tempo
1	1	2,3424	0,1843	4,1136	21	3	2,4211	0,1627	4,4571
1	2	0,7922	0,2413	7,6344	22	1	1,4136	0,2244	5,6297
1	3	3,1414	0,1693	2,7448	22	2	2,2044	0,2058	3,9801
2	1	0,4806	0,2656	8,8199	23	1	3,1388	0,1660	2,8055
2	2	2,7851	0,1717	3,4071	23	2	3,0462	0,1835	2,7010
2	3	3,5370	0,1587	2,1819	24	1	3,7287	0,1502	1,9536
3	1	2,4472	0,1826	3,9122	24	3	3,5555	0,1518	2,2457
3	2	4,9028	0,1345	0,1459	25	1	2,9364	0,1580	3,3680
3	3	2,5814	0,1943	3,4021	25	2	0,6548	0,2421	8,3961
4	1	1,7374	0,1993	5,3029	25	3	0,1592	0,3349	10,2913
4	2	3,4530	0,1935	1,9132	26	1	1,2820	0,2134	6,3788
4	3	2,8063	0,1833	3,1509	26	3	2,3712	0,1604	4,6502
5	2	2,1563	0,1939	4,3384	27	2	4,6494	0,1654	0,4396
5	3	3,8578	0,1497	1,7326	28	1	0,3430	0,2963	9,0424
6	1	0,3033	0,2319	12,0826	28	2	4,6970	0,1208	0,5176
6	3	0,9496	0,2116	7,8518	29	1	0,5806	0,2490	8,6490
7	1	2,5253	0,1675	4,0782	29	2	1,8313	0,1810	5,5480
7	3	1,3148	0,2358	5,6656	29	3	0,7386	0,2414	7,9219
8	1	0,9012	0,1457	11,7602	30	1	1,5835	0,1779	6,4638
8	2	1,5057	0,1702	7,0499	30	2	0,7153	0,2271	8,5622
8	3	1,0077	0,2136	7,4994	30	3	4,0332	0,1373	1,5650
9	1	0,4954	0,2648	8,7297	31	1	2,8913	0,1497	3,6602
9	2	1,3572	0,2082	6,2630	31	2	1,8604	0,1919	5,1520
9	3	4,5071	0,1492	0,6956	31	3	4,9802	0,1573	0,0252
10	1	1,3929	0,1905	6,7101	32	1	3,5986	0,1482	2,2193
10	2	4,3055	0,1330	1,1248	32	2	2,5279	0,2053	3,3224
10	3	1,0471	0,2111	7,4045	33	2	1,7403	0,1803	5,8553
11	1	1,3214	0,2157	6,1695	33	3	2,6197	0,1623	3,9819
11	3	0,5519	0,2374	9,2845	34	1	1,9658	0,1945	4,8003
12	1	1,5038	0,1918	6,2630	34	2	4,4890	0,1574	0,6847
12	2	1,4189	0,2316	5,4380	34	3	2,2095	0,1791	4,5601
14	2	1,3169	0,2065	6,4613	35	2	1,3175	0,2039	6,5395
15	1	4,3570	0,1729	0,7960	35	3	3,2694	0,1830	2,3221
15	3	4,0101	0,1556	1,4178	36	2	3,6838	0,1857	1,6450
16	1	4,9843	0,1623	0,0194	37	1	1,3116	0,1948	6,8712
16	2	4,1067	0,1579	1,2464	37	2	1,6851	0,1727	6,2977
17	1	1,2078	0,1838	7,7297	37	3	1,5514	0,2245	5,2141
17	2	1,3866	0,1746	7,3443	38	1	2,3824	0,2059	3,5998
17	3	4,0144	0,1828	1,2008	38	2	1,4348	0,1993	6,2654
18	1	3,7340	0,1340	2,1785	39	1	0,3917	0,2797	9,1058
18	2	1,0684	0,2260	6,8298	39	2	1,7549	0,2008	5,2138
18	3	1,1990	0,2150	6,6403	39	3	1,6987	0,2007	5,3782
19	1	2,7700	0,1714	3,4466	40	2	2,6882	0,1635	3,7949
19	2	4,1027	0,1743	1,1349	40	3	4,3144	0,1493	0,9880
19	3	2,6030	0,1747	3,7374	41	1	1,1228	0,2158	6,9208
20	1	1,1169	0,2256	6,6455	41	2	2,3682	0,1705	4,3841
20	2	4,1284	0,1563	1,2255	41	3	1,5940	0,2051	5,5747
20	3	2,1025	0,1710	5,0664	42	2	2,6237	0,1714	3,7619
21	1	4,7347	0,1325	0,4116	42	3	4,6991	0,1720	0,3608
21	2	0,3813	0,2928	8,7911					

Tabela 6 - Agrupamento dos acessos de tomateiro com o tempo (dias) até a ocorrência de 5% da doença, mostrando as falhas (1) e censuras (0) (FC) dos dados utilizados na análise de sobrevivência

Grupo	Acesso	Repetição	Tempo	FC	Grupo	Acesso	Repetição	Tempo	FC
1	3	1	3,9122	1	4	18	3	6,6403	0
1	3	2	0,1459	1	4	22	1	5,6297	0
1	3	3	3,4021	1	4	22	2	3,9801	1
1	19	1	3,4466	1	4	26	1	6,3788	0
1	19	2	1,1349	1	4	26	3	4,6502	1
1	19	3	3,7374	1	4	30	1	6,4638	0
1	23	1	2,8055	1	4	30	2	8,5622	0
1	23	2	2,701	1	4	30	3	1,565	1
1	31	1	3,6602	1	4	33	2	5,8553	0
1	31	2	5,152	1	4	33	3	3,9819	1
1	31	3	0,0252	1	4	35	2	6,5395	0
1	32	1	2,2193	1	4	35	3	2,3221	1
1	32	2	3,3224	1	4	38	1	3,5998	1
1	40	2	3,7949	1	4	38	2	6,2654	0
1	40	3	0,988	1	5	16	1	0,0194	1
2	4	1	5,3029	0	5	16	2	1,2464	1
2	4	2	1,9132	1	5	27	2	0,4396	1
2	4	3	3,1509	1	6	8	1	11,7602	0
2	20	1	6,6455	0	6	8	2	7,0499	0
2	20	2	1,2255	1	6	8	3	7,4994	0
2	20	3	5,0664	1	6	11	1	6,1695	0
2	21	1	0,4116	1	6	11	3	9,2845	0
2	21	2	8,7911	0	6	14	2	6,4613	0
2	21	3	4,4571	1	6	25	1	3,368	1
2	28	1	9,0424	0	6	25	2	8,3961	0
2	28	2	0,5176	1	6	25	3	10,2913	0
3	5	2	4,3384	1	6	29	1	8,649	0
3	5	3	1,7326	1	6	29	2	5,548	0
3	34	1	4,8003	1	6	29	3	7,9219	0
3	34	2	0,6847	1	6	39	1	9,1058	0
3	34	3	4,5601	1	6	39	2	5,2138	0
4	1	1	4,1136	1	6	39	3	5,3782	0
4	1	2	7,6344	0	7	24	1	1,9536	1
4	1	3	2,7448	1	7	24	3	2,2457	1
4	2	1	8,8199	0	7	36	2	1,6450	1
4	2	2	3,4071	1	7	42	2	3,7619	1
4	2	3	2,1819	1	7	42	3	0,3608	1
4	7	1	4,0782	1	8	12	1	6,263	0
4	7	3	5,6656	0	8	12	2	5,438	0
4	9	1	8,7297	0	8	37	1	6,8712	0
4	9	2	6,263	0	8	37	2	6,2977	0
4	9	3	0,6956	1	8	37	3	5,2141	0
4	10	1	6,7101	0	8	41	1	6,9208	0
4	10	2	1,1248	1	8	41	2	4,3841	1
4	10	3	7,4045	0	8	41	3	5,5747	0
4	17	1	7,7297	0	9	15	1	0,796	1
4	17	2	7,3443	0	9	15	3	1,4178	1
4	17	3	1,2008	1	10	6	1	12,0826	0
4	18	1	2,1785	1	10	6	3	7,8518	0
4	18	2	6,8298	0					

4.3. Análise de sobrevivência

Para o estudo da análise de sobrevivência, os grupos 3, 5, 7, 9 e 10 apresentaram baixos números de observações, dados respectivamente por 5, 3, 5, 2 e 2 (Tabela 6). Assim, seguindo recomendações de Allison (1995), esses valores foram desconsiderados na análise, pois, segundo esse autor, o baixo número de observações constitui um dos mais sérios fatores para a ineficiência de técnicas de análise de sobrevivência, principalmente em relação a aplicação de métodos probabilísticos, que exigem, em sua maioria, pressuposições assintóticas.

4.3.1. Estimador de Kaplan-Meier e Teste *Logrank*

As estimativas da função de sobrevivência, utilizando a técnica não-paramétrica de Kaplan-Meier, estão apresentadas na Tabela 7. Pode ser observado na Tabela 6 que, nos grupos 6 e 8, a maioria dos tempos foi de censura, ocorrendo apenas uma falha em cada um desses grupos. Em razão desse fato que na Tabela 7 foram estimadas apenas uma função de sobrevivência para os grupos 6 e 8.

Os acessos dos grupos 6 e 8 foram os que apresentaram maior sobrevivência, enquanto os do grupo 1, menor (Figura 8). Para se ter uma idéia da diferença entre esses grupos, basta verificar a probabilidade de severidade ser menor que 5%, por exemplo, num tempo de quatro dias. No grupo 1, essa probabilidade é menor que 0,0667, no grupo 2 é 0,54, no grupo 4 é 0,63 e no grupo 6 é 0,93, enquanto no grupo 8, 1,0.

Essas diferenças podem, ainda, ser observadas pelo teste *Logrank* (Tabela 8). Por este teste, pode-se concluir que a maioria das comparações duas a duas é significativa ($p < 0,05$), exceto, mas comparações dos grupos 2x4, 4x8 e 6x8, que não apresentaram diferenças interessantes.

Tabela 7 - Estimativas obtidas por Kaplan-Meier nos os grupos de acessos de tomateiro em que foram avaliados os tempos até 5% de severidade da requeima

Tempo	n	$\hat{S}(t)$	Erro-Padrão $\hat{S}(t)$	IC 95% $\hat{S}(t)$
Grupo-1				
0,0252	15	0,9333	0,0644	(0,8153 : 1,0000)
0,1459	14	0,8667	0,0878	(0,7106 : 1,0000)
0,9880	13	0,8000	0,1033	(0,6212 : 1,0000)
1,1349	12	0,7333	0,1142	(0,5405 : 0,9950)
2,2193	11	0,6667	0,1217	(0,4661 : 0,9530)
2,7010	10	0,6000	0,1265	(0,3969 : 0,9070)
2,8055	9	0,5333	0,1288	(0,3322 : 0,8560)
3,3224	8	0,4667	0,1288	(0,2717 : 0,8020)
3,4021	7	0,4000	0,1265	(0,2152 : 0,7430)
3,4466	6	0,3333	0,1217	(0,1630 : 0,6820)
3,6602	5	0,2667	0,1142	(0,1152 : 0,6170)
3,7374	4	0,2000	0,1033	(0,0727 : 0,5500)
3,7949	3	0,1333	0,0878	(0,0367 : 0,4840)
3,9122	2	0,0667	0,0644	(0,0100 : 0,4430)
5,1520	1	0	--	--
Grupo-2				
0,4120	11	0,9091	0,0867	(0,7540 : 1,0000)
0,5180	10	0,8182	0,1163	(0,6190 : 1,0000)
1,2260	9	0,7273	0,1343	(0,5060 : 1,0000)
1,9130	8	0,6364	0,1450	(0,4070 : 0,9950)
3,1510	7	0,5455	0,1501	(0,3180 : 0,9360)
4,4570	6	0,4545	0,1501	(0,2380 : 0,8680)
5,0660	5	0,3636	0,1450	(0,1660 : 0,7950)
Grupo-4				
0,6960	33	0,9697	0,0298	(0,9130 : 1,0000)
1,1250	32	0,9394	0,0415	(0,8610 : 1,0000)
1,2010	31	0,9091	0,0500	(0,8160 : 1,0000)
1,5650	30	0,8788	0,0568	(0,7740 : 0,9980)
2,1790	29	0,8485	0,0624	(0,7350 : 0,9800)
2,1820	28	0,8182	0,0671	(0,6970 : 0,9610)
2,3220	27	0,7879	0,0712	(0,6600 : 0,9400)
2,7450	26	0,7576	0,0746	(0,6250 : 0,9190)
3,4070	25	0,7273	0,0775	(0,5900 : 0,8960)
3,6000	24	0,6970	0,0800	(0,5570 : 0,8730)
3,9800	23	0,6667	0,0821	(0,5240 : 0,8490)
3,9820	22	0,6364	0,0837	(0,4920 : 0,8240)
4,0780	21	0,6061	0,0851	(0,4600 : 0,7980)
4,1140	20	0,5758	0,0860	(0,4300 : 0,7720)
4,6500	19	0,5455	0,0867	(0,3990 : 0,7450)
Grupo-6				
3,3680	15	0,9333	0,0644	(0,8153 : 1,0000)
Grupo-8				
4,3840	8	0,8750	0,1170	(0,6730 : 1,0000)

n: número de indivíduos sob risco e $\hat{S}(t)$: função de sobrevivência.

Observando-se a Figura 8 e a Tabela 8, era esperado não haver diferenças significativas entre os grupos 2x4 e 6x8, pois as curvas estão bem próximas. Entretanto, não era esperado que a comparação com 4x8 fosse não-significativa, uma vez que as curvas estão um pouco distantes. Uma possível explicação para este fato pode ser dada pela inconsistência do teste *Logrank* na presença de números de observações ($n=8$) relativamente baixos e desiguais, ou seja, tratamentos com diferentes números de repetições. Diante da constatação de problemas desta categoria, Silveira e Corrente (2004) propuseram a utilização do teste não-paramétrico de Friedman como alternativa ao *Logrank*.

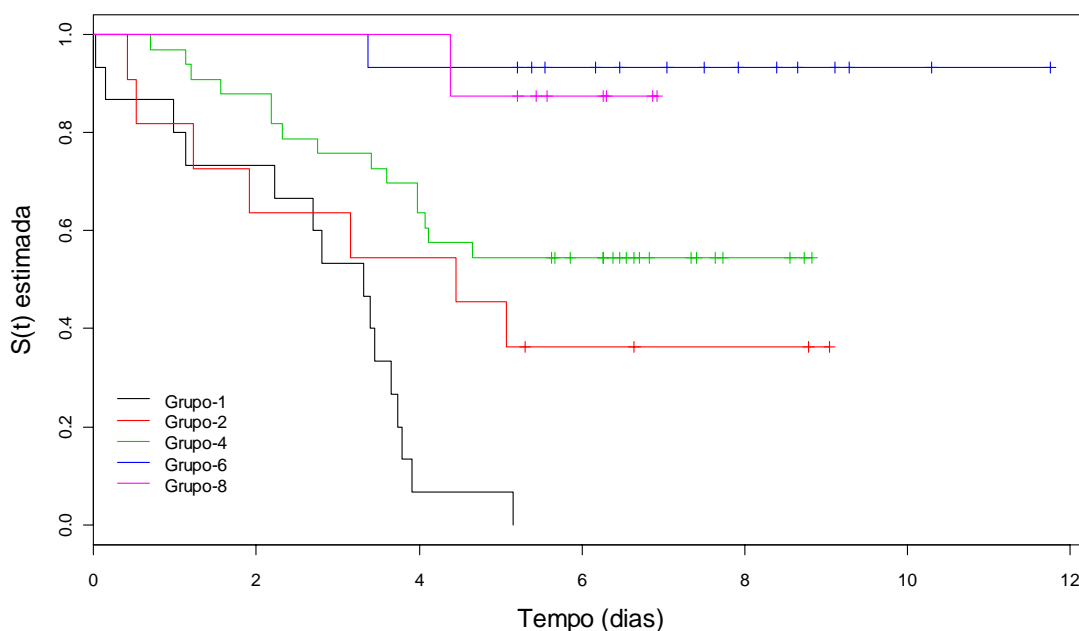


Figura 8 - Curvas de sobrevivência estimadas pelo método não-paramétrico de Kaplan-Meier para os tempos até 5% de severidade da requeima do tomateiro.

Tabela 8 - Valores de probabilidade (p-valor) referentes ao teste *Logrank*, considerando-se a hipótese de curvas de sobrevivência estimadas serem iguais

Comparação	p-valor	Comparação	p-valor	Comparação	p-valor
1 x 2	0,0391	2 x 4	0,2470	4 x 6	0,0122
1 x 4	< 0,001	2 x 6	0,0013	4 x 8	0,0941
1 x 6	< 0,001	2 x 8	0,0293	6 x 8	0,6700
1 x 8	< 0,001				

Fazendo uma comparação somente entre os grupos 1, 2, 4, 6 e 8, que foram comuns no estudo pelas técnicas de Kaplan-Meier e de ajuste de curvas de progresso da doença, pode-se verificar que os grupos 6, 8 e 4 foram os que obtiveram maior função de sobrevivência (Figura 8) e também os que sofreram menor progresso de doença (Figura 7). O grupo 2, por Kaplan-Meier, teve maior sobrevivência que o grupo 1, teste *Logrank* significativo (Tabela 8). Entretanto, pela técnica de ajuste de curvas de progresso da doença parece não ter havido diferenças entre as curvas desses grupos até aproximadamente 10 dias após a inoculação, mas depois desse tempo o grupo 1 apresentou menor progresso de doença que o grupo 2.

De forma geral, a discussão do parágrafo anterior revela certo antagonismo entre as técnicas de análise de sobrevivência e estudos de curvas de progresso de doença. Entretanto, esse antagonismo pode ser interpretado como fator importante em pesquisas na área de Fitopatologia, o qual está relacionado com o interesse do pesquisador no período de avaliação da doença. É possível entender que, estando o pesquisador interessado em avaliar todo o desenvolvimento da doença, mesmo em intervalos de tempo em que a enfermidade já expressou toda sua ação danosa contra a cultura, estudos de curvas de progresso mostram-se úteis, principalmente, para caracterizar estágios de maior ou menor infecção. Porém, se o interesse for estudar o efeito da doença até períodos de tempos considerados críticos, ou seja, acima dos quais o controle se torna inviável ou até mesmo impraticável, estudos de análise de sobrevivência são necessários, uma vez que permite classificar indivíduos (variedades, genótipos etc.) mais resistentes em relação a esse período crítico.

4.3.2. Modelos probabilísticos

As estimativas dos parâmetros dos modelos probabilísticos Exponencial, Weibull e Log-normal estão apresentados, respectivamente, nas Tabelas 9, 10 e 11 juntamente com os erros-padrão e Intervalos de Confiança (IC). Pode ser notado nessas tabelas, que os IC dos grupos 6 e 8 para as distribuições Exponencial e Weibull tiveram maior amplitude, fato este que pode ser explicado pelo grande número de observações censuradas, que propiciaram apenas um tempo de falha para cada um desses grupos.

Tabela 9 - Parâmetros do modelo exponencial estimados para o estudo dos tempos até 5% de severidade da requeima do tomateiro

Grupo	Parâmetro (α)	Erro-Padrão	IC 95%
1	2,6965	0,6962	1,6256 : 4,4728
2	6,6463	2,5121	3,1685 : 13,9413
4	11,1527	2,8796	6,7236 : 18,4994
6	112,0968	112,0967	15,7904 : 795,782
8	46,9636	46,9636	6,6155 : 333,3979

Tabela 10 - Parâmetros do modelo Weibull estimados para o estudo dos tempos até 5% de severidade da requeima do tomateiro

Grupo	Parâmetro (α)	Erro-Padrão	IC 95%	Parâmetro (γ)	Erro Padrão	IC 95%
1	2,8775	0,5418	1,9895 : 4,1617	1,4087	0,3321	0,8875 : 2,2360
2	6,8585	3,0486	2,8699 : 16,3905	0,8669	0,2866	0,4535 : 1,6572
4	10,2888	2,6913	6,1618 : 17,1800	1,1410	0,2675	0,7207 : 1,8064
6	75,5641	174,816	0,8111 : 7039,691	1,1741	1,1180	0,1816 : 7,5897
8	12,0303	8,605	2,961 : 48,8786	2,9744	2,7893	0,4734 : 18,6907

Tabela 11 - Parâmetros do modelo Log-normal estimados para o estudo dos tempos até 5% de severidade da requeima do tomateiro

Grupo	Parâmetro (μ)	Erro Padrão	IC 95%	Parâmetro (σ)	Erro Padrão	IC 95%
1	0,5298	0,365	-0,1855 : 1,2451	1,4135	0,2581	0,9883 : 2,0216
2	1,4112	0,4935	0,4438 : 2,3785	1,4871	0,4328	0,8406 : 2,6309
4	2,0042	0,2744	1,4664 : 2,5420	1,2133	0,2529	0,8064 : 1,8256
6	4,5772	2,5682	-0,4565 : 9,6108	1,7025	1,5218	0,2953 : 9,8167
8	2,4729	0,7397	1,0231 : 3,9226	0,5878	0,5151	0,1055 : 3,2748

Para comparar os modelos estimados, foi realizado o teste da razão de verossimilhança (TRV), comparando-se os modelos de interesse com o modelo Gama generalizado. Na Tabela 12 estão apresentados os resultados desse teste, na qual pode ser observado que, na maioria dos grupos, os modelos de interesse foram adequados, exceto no grupo 1; e o modelo Weibull, no grupo 4. Dessa forma, nas análises seguintes não foram apresentadas as estimativas para o grupo 1 e também para o modelo Weibull, no grupo 4.

Tabela 12 - Logaritmo natural da função $L(\theta)$ e resultado do Teste da Razão de Verossimilhança (TRV) dos modelos: exponencial, Weibull e Log-normal comparados com o modelo Gama generalizado

Modelo	$-\ln(L(\theta))$	TRV	Valor P	Modelo	$-\ln(L(\theta))$	TRV	Valor P
Grupo 1				Grupo 2			
Gama generalizado	25,03	-	-	Gama generalizado	18,70	-	-
Exponencial	29,88	9,70	0,0078	Exponencial	20,26	3,12	0,2101
Weibull	28,95	7,84	0,0051	Weibull	20,16	2,92	0,0875
Log-normal	34,42	18,78	0,0001	Log-normal	19,66	1,92	0,1659
Grupo 4				Grupo 6			
Gama generalizado	48,59	-	-	Gama generalizado	5,72	-	-
Exponencial	51,18	5,18	0,0750	Exponencial	5,72	0,00	1,0000
Weibull	51,02	4,86	0,0275	Weibull	5,71	0,02	0,8875
Log-normal	49,66	2,14	0,1435	Log-normal	5,61	0,22	0,6390
Grupo 8							
Gama generalizado	4,41	-	-				
Exponencial	4,85	0,88	0,6440				
Weibull	4,39	0,04	0,8415				
Log-normal	4,28	0,26	0,6101				

Na Tabela 13 são apresentadas as funções de sobrevivências estimadas por esses modelos. Com a finalidade de comparação, também nessa tabela é apresentada a função de sobrevivência estimada por Kaplan-Meier (Tabela 7).

Ainda na Tabela 13 pode ser notado, com relação aos grupos 2, 6 e 8, que as estimativas dos modelos probabilísticos em geral são maiores que as estimativas de Kaplan-Meier, principalmente nos tempos iniciais, em que o número de observações censuradas é grande. No grupo 4, as estimativas do modelo exponencial, nos tempos iniciais, apresentaram-se inferiores às de Kaplan-Meier e Log-normal.

Na Tabela 14 são apresentados os tempos medianos até a ocorrência de 5% da severidade da doença nas plantas de tomateiro. Pode ser notado, nessa tabela, que o estimador de Kaplan-Meier e os modelos Exponencial, Weibull e Log-normal fornecem, para os grupos 1 e 2, o número de dias como estimativa do tempo mediano de severidade da doença das plantas, ou seja, esses são os tempos em que 50% das plantas permanecem com a severidade da doença abaixo de 5%. Nos grupos 4, 6 e 8 não houve uma estimativa adequada, pois as plantas com os maiores tempos observados foram censuradas. Os modelos probabilísticos forneceram estimativas superiores à estimativa de Kaplan-Meier.

Tabela 13 - Estimativa da função de sobrevivência nos tempos até 5% de severidade da requeima em tomateiro usando-se o estimador de Kaplan-Meier e os modelos exponencial, Weibull e Log-normal

Tempo	Kaplan-Meier	Expon.	Weibull	Log-normal	Tempo	Kaplan-Meier	Expon.	Log-normal
Grupo 2					Grupo 4			
0,4116	0,9091	0,9399	0,9164	0,9389	0,6956	0,9697	0,9395	0,9745
0,5176	0,8182	0,9251	0,8990	0,9180	1,1248	0,9394	0,9041	0,9400
1,2255	0,7273	0,8316	0,7987	0,7917	1,2008	0,9091	0,8979	0,9333
1,9132	0,6364	0,7499	0,7185	0,6959	1,5650	0,8788	0,8691	0,9002
3,1509	0,5455	0,6225	0,6008	0,5703	2,1785	0,8485	0,8226	0,8438
4,4571	0,4545	0,5114	0,5025	0,4777	2,1819	0,8182	0,8223	0,8435
5,0664	0,3636	0,4666	0,4634	0,4435	2,3221	0,7879	0,8120	0,8308
5,3029	0,3636	0,4503	0,4493	0,4314	2,7448	0,7576	0,7818	0,7938
6,6455	0,3636	0,3679	0,3779	0,3727	3,4071	0,7273	0,7368	0,7394
8,7911	0,3636	0,2664	0,2893	0,3040	3,5998	0,6970	0,7241	0,7245
9,0424	0,3636	0,2565	0,2806	0,2975	3,9801	0,6667	0,6999	0,6962
Grupo 6					3,9819	0,6364	0,6997	0,6960
3,3680	0,9333	0,9704	0,9744	0,9759	4,0782	0,6061	0,6937	0,6891
5,2138	0,9333	0,9546	0,9576	0,9572	4,1136	0,5758	0,6915	0,6866
5,3782	0,9333	0,9532	0,9561	0,9555	4,6502	0,5455	0,6590	0,6499
5,5480	0,9333	0,9517	0,9545	0,9537	5,6297	0,5455	0,6036	0,5900
6,1695	0,9333	0,9464	0,9486	0,9474	5,6656	0,5455	0,6017	0,5880
6,4613	0,9333	0,9440	0,9458	0,9444	5,8553	0,5455	0,5915	0,5774
7,0499	0,9333	0,9390	0,9401	0,9384	6,2630	0,5455	0,5703	0,5556
7,4994	0,9333	0,9353	0,9358	0,9338	6,2654	0,5455	0,5702	0,5554
7,9219	0,9333	0,9318	0,9317	0,9296	6,3788	0,5455	0,5644	0,5496
8,3961	0,9333	0,9278	0,9270	0,9249	6,4638	0,5455	0,5601	0,5453
8,6490	0,9333	0,9257	0,9245	0,9224	6,5395	0,5455	0,5563	0,5415
9,1058	0,9333	0,9220	0,9200	0,9179	6,6403	0,5455	0,5513	0,5365
9,2845	0,9333	0,9205	0,9182	0,9161	6,7101	0,5455	0,5479	0,5330
10,2913	0,9333	0,9123	0,9082	0,9064	6,8298	0,5455	0,5421	0,5272
11,7602	0,9333	0,9004	0,8935	0,8927	7,3443	0,5455	0,5176	0,5034
Grupo 8					7,4045	0,5455	0,5148	0,5007
4,3841	0,8750	0,9109	0,9516	0,9547	7,6344	0,5455	0,5043	0,4906
5,2141	0,8750	0,8949	0,9202	0,9189	7,7297	0,5455	0,5000	0,4866
5,4380	0,8750	0,8907	0,9101	0,9076	8,5622	0,5455	0,4641	0,4530
5,5747	0,8750	0,8881	0,9035	0,9004	8,7297	0,5455	0,4571	0,4467
6,2630	0,8750	0,8752	0,8663	0,8612	8,8199	0,5455	0,4535	0,4434
6,2977	0,8750	0,8745	0,8643	0,8591				
6,8712	0,8750	0,8639	0,8278	0,8233				
6,9208	0,8750	0,8630	0,8244	0,8201				

Tabela 14 - Tempo mediano até a ocorrência de 5% da severidade da requeima em plantas de tomateiro

Grupo	Kaplan-Meier	Exponencial	Weibull	Log-normal
1	3,1	na ^{1/}	na	na
2	3,8	4,6	4,5	4,2
4	*	*	na	*
6	*	*	*	*
8	*	*	*	*

1/ na: modelo não adequado e *: estimativa não adequada.

Na Figura 9 são apresentadas as curvas de sobrevivências estimadas pelos modelos Exponencial, Weibull e Log-normal *versus* a curva de sobrevivência estimada por Kaplan-Meier. Pode ser observado, nos grupos 2 e 4, que os modelos paramétricos apresentaram ajustes satisfatórios. Entretanto, a ocorrência de apenas uma falha nos grupos 6 e 8 fizeram que a curva estimada por Kaplan-Meier apresentasse apenas um degrau, tornando essa comparação difícil de ser verificada.

4.3.3. Comparação dos parâmetros dos modelos probabilísticos

Nas Figuras 10, 11, 12, 13 e 14 são apresentados os intervalos de confiança (IC) de 95% para os parâmetros dos modelos probabilísticos. Na Figura 10 pode ser observado que o limite inferior do IC para o modelo exponencial do grupo 6 foi maior que o limite superior do grupo 2, ou seja, não houve sobreposição do IC. Isso indica que houve diferença significativa no nível de 5% de probabilidade entre os parâmetros desse modelo nesses grupos. Nas demais comparações houve sobreposição dos IC, indicando que não ocorreram diferenças significativas entre os parâmetros do mesmo modelo.

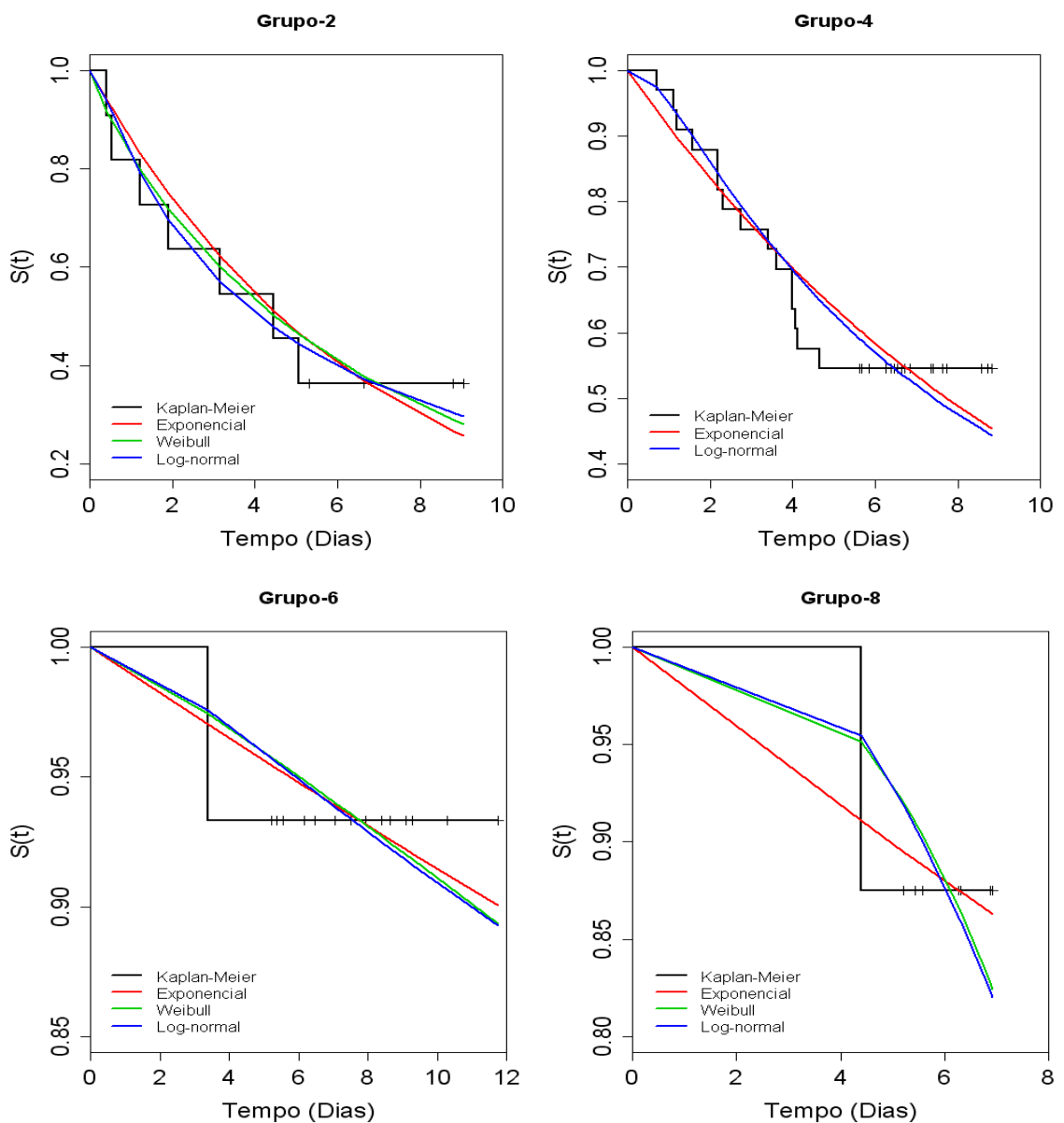


Figura 9 - Curvas de sobrevivência nos os tempos até 5% de severidade da requeima em tomateiro, estimadas pelos modelos Exponencial, Weibull e Log-normal *versus* a curva de sobrevivência estimada por Kaplan-Meier.

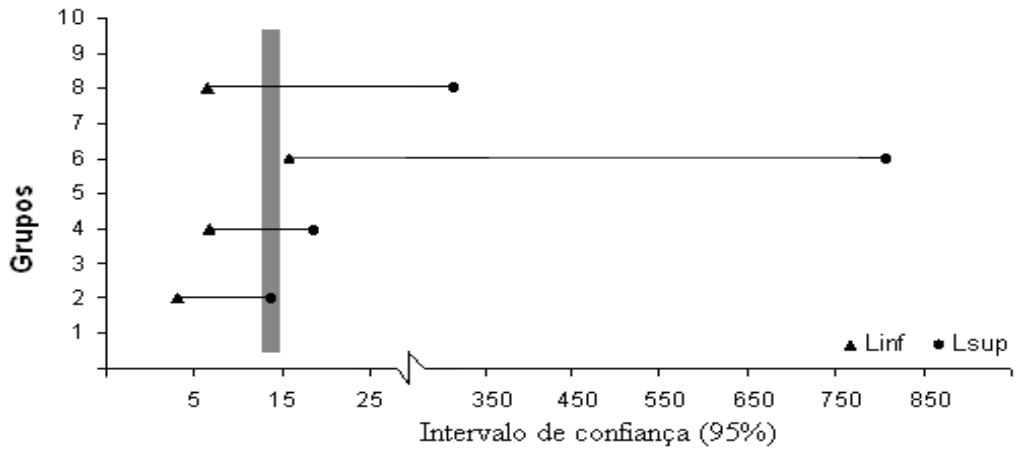


Figura 10 - Intervalo de confiança (95%) para o parâmetro α do modelo Exponencial.

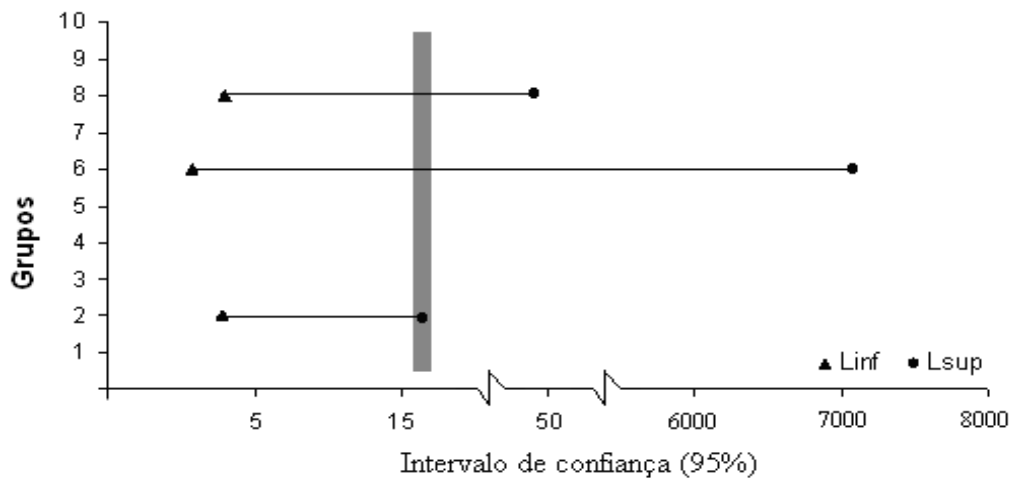


Figura 11 - Intervalo de confiança (95%) para o parâmetro α do modelo Weibull.

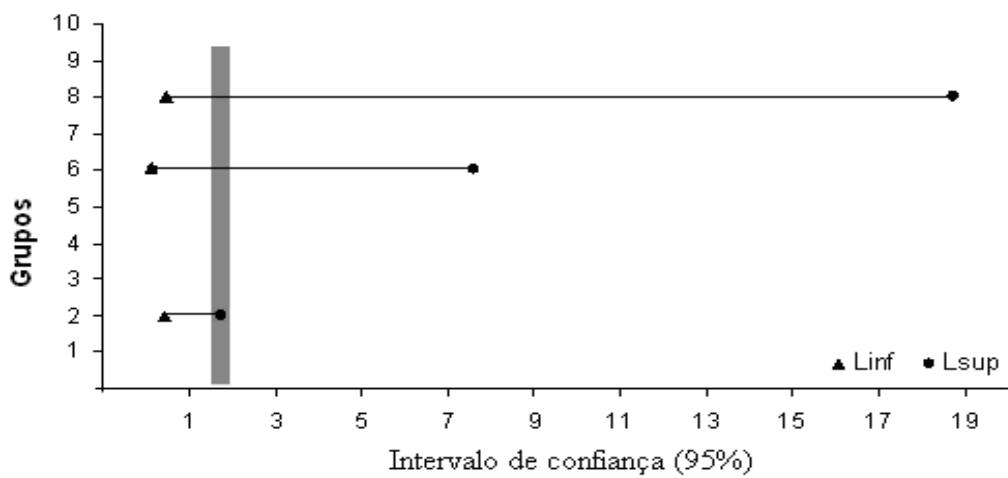


Figura 12 - Intervalo de confiança (95%) para o parâmetro γ do modelo Weibull.

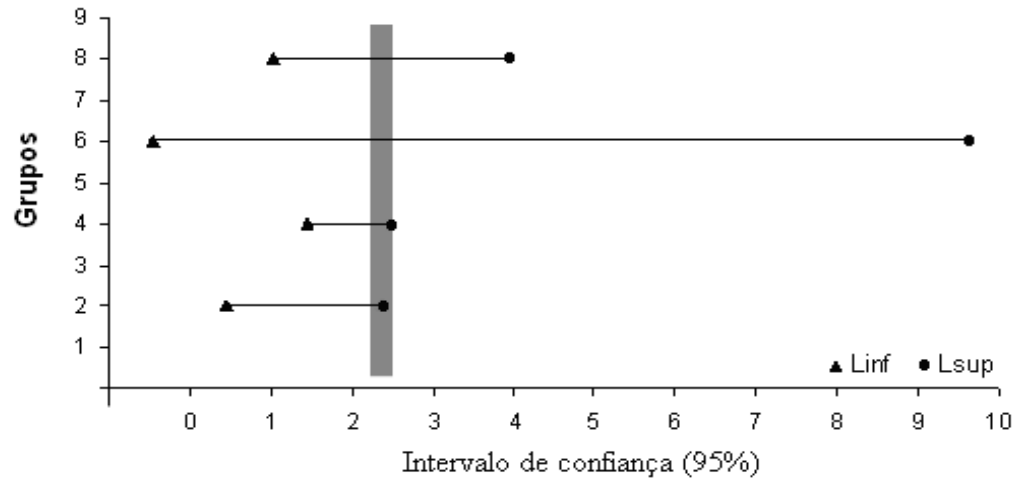


Figura 13 - Intervalo de confiança (95%) para o parâmetro μ do modelo Log-normal.

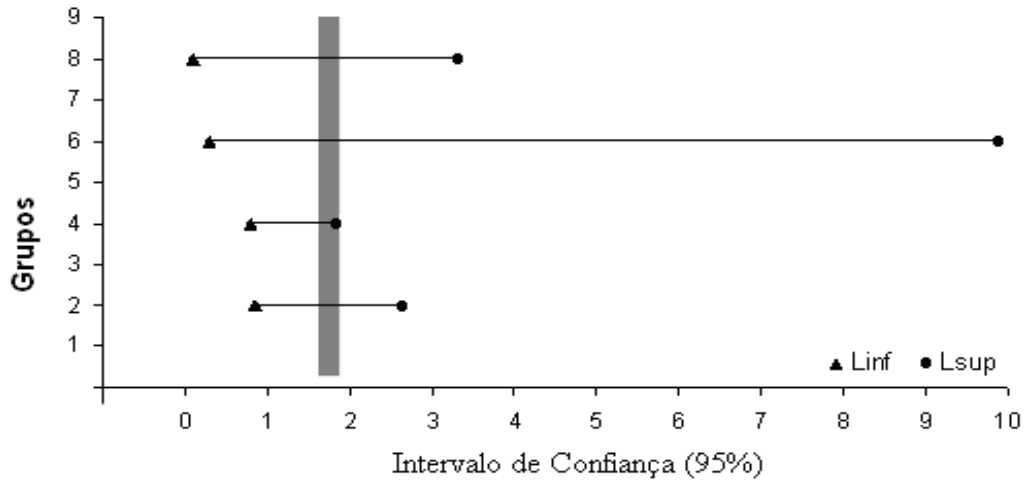


Figura 14 - Intervalo de confiança (95%) para o parâmetro σ do modelo Log-normal.

5. CONCLUSÕES

Diante dos problemas que a requeima causada por *Phytophthora infestans* pode provocar às lavouras de tomate, neste trabalho foram ajustados modelos para descrever o progresso da doença, bem como foram empregadas algumas técnicas de análise de sobrevivência no tempo até 5% de severidade da doença nas plantas.

Utilizando um conjunto de dados reais sobre essa doença, este estudo permitiu concluir que:

- Foi possível ajustar o modelo exponencial para descrever o progresso da requeima do tomateiro e agrupar os acessos estudados em 10 grupos. O acesso BGH-6 obteve um progresso de doença menor que os demais, caracterizando-se, assim, sua maior resistência à doença.
- A regressão inversa possibilitou estimar o tempo até a ocorrência de 5% da severidade da requeima do tomateiro.
- Pela técnica não-paramétrica, de Kaplan-Meier, foi possível estimar as curvas de sobrevivência dos acessos de tomateiro pertencentes aos grupos 1, 2, 4, 6 e 8. Utilizando o teste *Logrank*, pode-se concluir que a maioria das comparações duas a duas foi significativa ($p < 0,05$) exceto nas comparações dos grupos 2x4, 4x8 e 6x8.

- O uso dos modelos probabilísticos Exponencial, Weibull e Log-normal possibilitou a estimação das curvas de sobrevivência dos grupos 2, 4, 6 e 8, exceto o grupo 4, no qual o modelo Weibull não foi adequado. Comparando os modelos probabilísticos com a técnica não-paramétrica, as curvas desses modelos nos grupos 2 e 4 apresentaram ajustes satisfatórios em relação à curva estimada por Kaplan-Meier.

6. REFERÊNCIAS

AALEN, O.O. Two examples of modelling heterogeneity in survival analysis. **Scandinavian Journal of Statistics**, v. 14, p. 19-25, 1987.

ALLISON, P.D. **Survival analysis using the SAS system**. A practical guide. Cary, NC: SAS Institute, 1995.

BERGAMIN FILHO, A. Curvas de progresso da doença. In: BERGAMIN FILHO, A.; KIMATI, H. **Manual de fitopatologia**. São Paulo: Agronômica Ceres, 1995. v.1, p. 602-626.

BLANCARD, D. **Enfermedades del tomate**. Madrid: Ediciones MundiPrensa, 1996. 212 p.

BRANNEN, P.M.; SCHERM, H.; BRUORTON, M.D. Fungicidal control of septoria leaf spot of blueberry, 2002. **Fung. Nemat. Tests**, v. 58, p. SMF019, 2003.

BOFF, P. **Epidemiologia e controle químico da mancha de estenfílio (*Stemphylium solani*) e da pinta preta (*Alternaria solani*) em dois sistemas de condução do tomateiro**. Viçosa, MG: 1988. 192 f. Dissertação (Mestrado em Fitotecnia) – Universidade Federal de Viçosa, Viçosa.

BOLFARINE, H.; RODRIGUES, J.; ACHCAR, J.A. **Análise de sobrevivência**. Rio de Janeiro: 2ª Escola Nacional de Modelos de Regressão, 1991.

CAMARGO FILHO, W.P.; MAZZEI, A.R. Mercado mundial de tomate e o MERCOSUL. **Informações Econômicas**, v. 27, n.10, p. 25-38, 1997.

CAMPBELL, C.L.; MADDEN, L.V. **Introduction to plant disease epidemiology**. New York: Wiley-Interscience Publication; John Wiley & Sons, Inc., 1990.

CARVALHO, M.S.; ANDREOOZI, V.L.; CODEÇO, C.T.; BARBOSA, M.T.S.; SHIMAKURA, S.E. **Análise de sobrevida: teoria e aplicações em saúde**. Rio de Janeiro: Editora Fiocruz, 2005.

CHALITA, L.V.A.S.; SILVEIRA, L.V.A.; COLOSIMO, E.A.; PASSOS, J.R.S. Modeling grouped survival data with time-dependent covariates. **Communications in Statistics-Simulation And Computation**, Hamilton, Canadá, v. 35, n. 4, p. 975-981, 2006.

CHIANG, C.L. **Introduction to stochastic process in biostatistics**. New York: John Wiley & Sons, 1968.

COLLET, D. **Modelling survival data in medical research**. 1. ed. London: Chapman and Hall, 1994.

COLOSIMO, E.A. Análise de Sobrevivência Aplicada. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA (RBRAS), 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA E EXPERIMENTAÇÃO AGRONÔMICA (SEAGRO), 9., 1995, Piracicaba. Resumos... Piracicaba, SP: ESALQ, 1995.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blucher, 2006.

COX, D.R.; HINKLEY, D.V. **Theoretical statistics**. London: Chapman & Hall, 1974.

CROW, E.W.; SHIMIZU, K. **Lognormal distributions**. New York: Marcel Dekker, 1988.

DEHNE, H.W.; OERKE, E.C. Impacts of diseases and disease control on crop production. In: HUTSON, D.; MYAMAMOTO, J. (Eds.). **Fungicide activity**. Hamburg: John Wiley & Sons, 1998. p. 1-21.

DUARTE, H.S.S.; ZAMBOLIM, L.; JESUS JUNIOR, W.C. Manejo da requeima do tomateiro industrial empregando sistema de previsão. **Summa Phytopathologica**, v. 33, n. 4, p. 328-334, 2007.

FLETCHER, R.H.; FLETCHER, S.W.; WAGNER, E.H. **Epidemiologia clínica: elementos Essenciais**. 3. ed. Porto Alegre: Artes Médicas, 1996.

FREITAS, M.A.; COLOSIMO, E.A. **Confiabilidade**: análise de tempo de falha e testes acelerados. Belo Horizonte: Fundação Cristiano Ottoni, 1997.

GEHAN, E.A. A generalized wilcoxon test for comparing arbitrarily singly-censored samples. **Biometrika**, v. 52, p. 203-223, 1965.

JOHNSON, N.L.; KOTZ, S. **Distributions in statistics: continuous distributions**. New York: John Wiley, 1970.

JONES, J.B.; STALL, R.E.; ZITTER, T.A. **Compendium of tomato diseases**. St Paul: APS, 1993. 73 p.

KAHN, H.A.; SEMPOS, C.T. **Statistical methods in epidemiology**. New York: Oxford University Press, 1989.

KALBFLEISCH, J.D., PRENTICE, R.L. **The statistical analysis of failure time data**. New York: John Wiley & Sons, 1980.

KAPLAN, E.L.; MEIER, P. Nonparametric estimation from incomplete observation. **Journal of the American Statistical Association**, v. 53, p. 457-481, 1958.

KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**. New York: BBU Press and John Wiley Sons Inc., 2000. 574 p.

KIM, K. O.; KUO, W. Percentile life and reability as performance measures in optimal system design. **International Journal of Reability, Quality and Safety Engineering**, v. 35, p. 1133-1142, 2003.

KLEIN, J.P.; MOESCHBERGER, M.L. **Survival analysis: thechniques for censored and truncated data**. New York: Springer-Verlag, 1997.

KUROZAWA, C.; PAVAN, M.A. Doenças do tomateiro (*Lycopersicon esculentum*). In: KIMATI, H. (Ed.). **Manual de fitopatologia**. São Paulo: Agronômica Ceres, 1997. v. 2, p. 690-719.

LABORY, C.R.G.; SANTA-CECÍLIA, L.V.C.; MALUF, W.R.; CARDOSO, M.G.; BEARZOTTI, E.; SOUZA, J.C. Seleção indireta para teores de 2-tridecanona em tomateiros segregantes e sua relação com a resistência à traça do tomateiro. **Pesquisa Agropecuária Brasileira**, v. 34, p. 723-739, 1999.

LAI, C.D.; XIE, M.; MURTHY, D.N.P. A modified Weibull distribution. **International Journal of Reability, Quality and Safety Engineering**, v. 52, n.1, p.33-37, 2003.

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

LEE, E.T. **Statistical methods for survival data analysis**. 2nd ed. New York: John Wiley & Sons, 1992.

LOPES, C.A.; SANTOS, J.R.M. **Doenças do tomateiro**. Brasília: EMBRAPA/CNPH, 1994. 67 p.

LUTHIANI, C.P.; JESUS JUNIOR, W.C.; BOGO, A.; LAZAROTO, A.; SILVA, A.; AMARANTE, C.V.T. Análise temporal da sarna da macieira nas cultivares Royal Gala e Fuji sob os sistemas convencional e orgânico de produção. **Fitopatologia brasileira**, v. 31, n. 6, p. 585-591, 2006.

MANTEL, N.; HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. **J. Nat. Cancer Ins.**, v. 22, p. 719-748, 1959.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Rep.**, v. 50, p.163-170, 1966.

MARUBINI, E.; VALSECHI M.G. **Analysing survival data from clinical trials and observational studies**. Chichester (England): Wiley-Liss, 1995.

MOTULSKY, H.J. **Intuitive biostatistics**. Oxford University Press (Oxford), 1995. Disponível em: <<http://www.graphpad.com/www/book/book.htm>>. Acesso em: 12 setembro 2008.

MUDHOLKAR, G. D. S.; SRIVASTAVA, D. K. Exponentiated Weibull family for analyzing bathtub failure-rate data. **International Journal of Reability, Quality and Safety Engineering**, v. 42, n. 2, p. 299-302, 1993.

NELSON, W. **Accelerated life testing**: statistical models, data analysis and test plans. New York: John Wiley & Sons, 1990.

PAULA, A.R. **Seleção de isolados de fungos entomopatogênicos, visando o controle de adultos do mosquito *Aedes aegypti* (Díptera: Culicidae)**. Campos dos Goytacases, RJ: UENF, 2007. Tese (Doutorado) – Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacases.

PEREIRA, E.F.; LOPES, J.R.S.; TURATI, D.T.; MUNHOZ, C.; CORRENTE, J.E. Influência das condições hídricas do solo e da temperatura na sobrevivência e alimentação de *Oncometopia facialis* (Hemiptera: *Cicadellinae*) em seedlings de citros. **Arquivos do Instituto Biológico**, v. 72, n. 3, p. 343-351, 2005.

PIRES, M.C.; COLOSIMO, E.A. Identificação dos fatores determinantes de crescimento em crianças com baixa estatura: uma aplicação do modelo de Cox estratificado combinado com um modelo para dados longitudinais. **Revista Brasileira de Estatística**, v. 66, p. 51-61, 2005.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2008. Disponível em: <URL <http://www.R-project.org>>.

RIBEIRO, B.N.; SILVA, F.F.; RODRIGUES, G.S.; SILVA, D.J.H.; MIZUBUTI, E.S.G.; CECON, P.R. Comparação de curvas de progresso da requeima do tomateiro entre acessos do banco de germoplasma de hortaliças - UFV sob um enfoque multivariado. In: RBRAS, 52.; SEAGRO, 12., 2007, Santa Maria. **Anais...** Santa Maria, RS, 2007.

REZENDE, D.M.L.C.; MUNIZ, J.A.; FERREIRA, D.F.; SILVA, F. F.; AQUINO, L.H. Ajuste de modelos de platô de resposta para a exigência de zinco em frangos de corte. **Ciência e Agrotecnologia**, v. 31, p. 468-478, 2007.

SAS. Statistical Analysis System Institute – SAS/STAT. **User's guide** – Version 6.11. 4th. ed. Cary, 1996. v. 2, 842 p.

SILVA, J.B.C.; GIORDANO, L.B. Produção mundial e nacional. In: SILVA, J.B.C.; GIORDANO, L.B. (Eds.). **Tomate para processamento industrial**. Brasília: EMBRAPA Hortaliças, 2000. p. 8-11.

SCHERM, H.; OJIAMBO, P.S. Applications of survival analysis in botanical epidemiology. **Phytopathology**, v. 94, n. 9, p. 1022-1026, 2004.

SILVEIRA, L.V.A.; CORRENTE, J.E. Aplicação do teste de Friedman na comparação de curvas de sobrevivência. In: REUNIÃO ANUAL DA RBRAS, 49., 2004, Uberlândia. **Programa e Resumos...** Uberlândia, MG, 2004. p.113-113.

STACY, E.W. A generalization of the Gamma distribution. **Ann. Math. Stat.**, v. 33, p. 1187-1192, 1962.

STEVENSON, W.R. An integrated program for managing potato late blight. **Plant Disease**, v. 67, p.1047-1048, 1983.

SOUZA, G.S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: Empresa Brasileira de Pesquisa Agropecuária/SPI; Empresa Brasileira de Pesquisa Agropecuária/SEA, 1998. 489 p.

SUINAGA, F.A.; CASALI, V.W.D.; SILVA, D.J.H.; PICANÇO, M.C. Dissimilaridade genética de fontes de resistência de *Lycopersicon* spp. a *Tuta absoluta* (Meyrick, 1917) (Lepidoptera: Gelechiidae). **Revista Brasileira de Agrociência**, v. 9, n. 4, p. 371-376, 2003.

TEIXEIRA, M.T.B.; FAERSTEIN, E.; LATORRE, M.R. Tendências de análise de sobrevivência. **Cadernos de Saúde Pública**, maio-jun. 2002.

TÖFOLI, J.G.; DOMINGUES, R.J.; GARCIA JUNIOR, O. Controle da requeima do tomateiro com fungicidas e seus reflexos na produção. **Arq. Inst. Biol.**, v. 70, n. 4, p. 473-482, 2003.

XIE, M.; GAUDO, O.; BRACQUEMOND, C. Redefining failure rate function for discrete distributions. **International Journal of Reability, Quality and Safety Engineering**, v. 9, n. 3, p.275-285, 2002.

WEIBULL, W. A statistical theory of the strength of materials. **Ingeniors Vetenskaps Akademien**, v. 151, 1939.