

**ESTUDO DA VIABILIDADE DA APLICAÇÃO
DA ANÁLISE DE SOBREVIVÊNCIA EM
DADOS ENTOMOLÓGICOS**

PAULO JOSÉ PEREIRA

2002

2
PAULO JOSÉ PEREIRA

**ESTUDO DA VIABILIDADE DA APLICAÇÃO DA ANÁLISE DE
SOBREVIVÊNCIA EM DADOS ENTOMOLÓGICOS**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

Orientador

Prof. Mário Javier Ferruz Vivanco

LAVRAS
MINAS GERAIS – BRASIL
2002

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Pereira, Paulo José

Estudo da viabilidade da aplicação da análise de sobrevivência em dados entomológicos / Paulo José Pereira. -- Lavras : UFLA, 2002.

81 p. : il.

Orientador: Mário Javier Ferrua Vivanco.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Análise de sobrevivência. 2. Censura. 3. Tempo de vida. I. Universidade Federal de Lavras. II. Título.

CDD-519.54
-630.2195

PAULO JOSÉ PEREIRA

**ESTUDO DA VIABILIDADE DA APLICAÇÃO DA ANÁLISE DE
SOBREVIVÊNCIA EM DADOS ENTOMOLÓGICOS**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, Área de Concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

APROVADA em 22 de fevereiro de 2002

Pesq. Edwin Moises Marcos Ortega

ACNIELSEN

Prof. Lucas Monteiro Chaves

DEX - UFLA


Prof. Mário Javier Ferrua Vivanco
DEX - UFLA
(Orientador)

**LAVRAS
MINAS GERAIS - BRASIL**

DEDICO

Ao meu pai, Carlos e minha mãe, Francisca, pelo apoio, incentivo, amor eterno e por todos os sacrifícios que fizeram para que eu pudesse atingir meus objetivos. A eles devo tudo.

Aos meus irmãos, Sílvia, Claudinei, Luci Mara e Francisco, que sempre me ajudaram em tudo, sempre me incentivaram e por serem grandes amigos.

A Monica, por estar, mesmo distante, sempre ao meu lado, nos momentos mais difíceis e por seu amor, que é a coisa mais preciosa da minha vida.

AGRADECIMENTOS

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade de realizar este curso.

Ao CNPq, pela concessão da bolsa de estudos.

Ao Prof. Doutor Mário Javier Ferrua Vivanco, pela orientação, amizade, compreensão e por sempre confiar na minha capacidade para a realização deste trabalho.

Aos professores do DEX, Augusto, Daniel, Delly, Eduardo, Joel, Júlio, Lucas, Luiz Henrique, Marcelo, Thelma, pelos ensinamentos e, principalmente, pela disposição em passar conhecimentos.

Às funcionárias do DEX, Andréia, Edila, Ester, Maria, Maristela e Rogéria, pela simpatia e imensa disposição em nos atender.

Aos meus colegas de curso, Ceile, Chico, Douglas, Ednaldo, Flávio, Ivana, José Marcelo, Lívia, Marcelo, Marcos, Paulo César e Sérgio, pela colaboração, companheirismo e amizade que foram fundamentais durante o curso.

Ao Cirilo, pela colaboração no desenvolvimento computacional deste trabalho.

Aos amigos Adriano, Andréa e Andréa Leonel, pela força que me deram antes e durante este curso.

A Deus, por ter me oferecido esta oportunidade e dado as forças necessárias para que este sonho tenha se realizado.

Enfim, a todos que contribuíram para a realização deste trabalho.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT.....	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	4
2.1 Inferência não-paramétrica.....	10
2.1.1 Estimador de Kaplan-Meier.....	11
2.2 Inferência paramétrica.....	14
2.2.1 Distribuição exponencial.....	14
2.2.2 Distribuição de Weibull.....	15
2.2.3 Distribuição log-normal.....	17
2.2.4 Distribuição gama.....	18
2.2.5 Distribuição gama generalizada.....	19
2.3 Comparação de duas curvas de sobrevivência.....	20
2.3.1 Teste de Gehan	21
2.3.2 Teste F de Cox.....	21
2.4 Modelos de regressão.....	23
2.4.1 Modelos de locação e escala.....	23
2.4.1.1 Modelo de regressão exponencial.....	24
2.4.1.2 Modelo de regressão Weibull.....	25
2.4.1.3 Modelo de regressão log-normal.....	25
3. MATERIAL E MÉTODOS.....	27
3.1 Materiais.....	28
3.2 Determinação do modelo paramétrico adequado.....	30
3.2.1 Gráficos de linearização da função de sobrevivência.....	30
3.3 Máxima verossimilhança.....	31

3.3.1 Máxima verossimilhança com a presença de censura.....	32
3.4 Simulação das amostras.....	35
3.5 Intervalos de confiança e teste de hipóteses para comparação de médias de estimadores.....	36
3.6 Teste log-rank.....	39
3.7 Fator de correção para estimativas censuradas.....	43
4 RESULTADOS E DISCUSSÃO.....	45
4.1 Adequação do modelo.....	45
4.2 Intervalo de confiança e teste de comparação de médias de estimadores.....	47
4.3 Teste log-rank.....	52
4.4 Fator de correção para estimativas censuradas.....	54
5. CONCLUSÕES.....	57
REFERÊNCIAS BIBLIOGRÁFICAS	59
ANEXOS.....	63

RESUMO

Pereira, Paulo José. **Estudo da viabilidade da aplicação da análise de sobrevivência em dados entomológicos**. Lavras: UFLA, 2002. 81p. (Dissertação – Mestrado em Agronomia/Estatística e Experimentação Agropecuária)¹.

Para realizar uma análise correta e confiável de um experimento que estuda tempos de vida na presença de dados censurados (informação incompleta) é fundamental a utilização das técnicas de Análise de Sobrevivência. Apesar da grande quantidade de pesquisas e trabalhos publicados sobre o tempo de vida na área da entomologia, não é comum encontrar algum que tenha utilizado estas técnicas. Isto provavelmente decorre da falta de conhecimento sobre este método, que passou a ser utilizado com frequência em diversas áreas, inclusive na agropecuária. Além de mostrar as técnicas de Análise de Sobrevivência, este trabalho tem também como objetivo estimar e comparar curvas de sobrevivência, para avaliar a possibilidade de aplicar mecanismos de censura, tanto a aleatória como a do tipo I, em dados entomológicos. Esses mecanismos atualmente são aplicados implicitamente sem o tratamento dos dados na análise via Análise de Sobrevivência. Outro objetivo é estudar a viabilidade da aplicação do teste log-rank na comparação de duas curvas de sobrevivência, à medida que se introduz uma porcentagem maior de censura nas amostras. Para realizar este estudo, foram utilizados dados reais entomológicos de machos e fêmeas da *Ceraeochrysa cubana*, sem censura e, a partir deles montou-se um sub-rotina em SAS®. Nessa sub-rotina, por meio de simulação, introduziram-se as censuras aleatória e do tipo I, aplicando assim o teste log-rank para comparação das curvas resultantes, isto para diferentes porcentagens de censura. Utilizando-se os dados reais, o teste mostrou que, para os tempos de vida sem censura de machos e fêmeas, as curvas de sobrevivência eram diferentes e, à medida que se introduziu a censura tipo I, elas se tornaram semelhantes, a partir de 30% de censura. Foi detectado, por meio de simulação, que com um aumento da presença de censura, a precisão do teste log-rank foi diminuindo. Por meio dos conceitos do teorema central do limite, foi possível determinar que, mesmo com a introdução de censura, as funções de sobrevivência de machos e fêmeas continuaram sendo diferentes. Além disso, a introdução de censura aos dados fez com que suas curvas de sobrevivência se tornassem diferentes das respectivas curvas com os tempos de vida cuja informação foi completa. Além dos resultados mostrados, esse trabalho ainda propõe uma forma de determinar um fator de correção dos estimadores para amostras com dados censurados, para aproximá-los dos estimadores para as amostras sem a presença de censura.

¹ Orientador: Mário Javier Ferrua Vivanco – UFLA

ABSTRACT

Pereira, Paulo José. **Study of the Viability of the application of Survival Analysis in Entomological Data**. Lavras: UFLA , 2002 . 81p. (Dissertation – Master in Agronomy / Statistics and Agricultural Experimentation)¹

To accomplish a correct and reliable analysis of an experiment which studies lifetimes in the presence of censored data (incomplete information) , the use of the techniques of Survival Analysis is fundamental . In spite of the great amount of research and works published about the lifetime in the area of entomology , it is not common to find some which had utilized these techniques. This probably occurs , due to the lack of knowledge on this method , which began to be utilized with frequency in a number of areas , inclusive in agriculture and animal husbandry . In addition to showing the techniques of Survival Analysis , this work is also concerned with estimating and comparing survival curves to evaluate the possibility of applying censoring mechanisms both random and type I in entomological data ,these mechanisms at present being applied implicitly without the treatment of the data in the analysis via Survival Analysis. Another objective is to study the viability of the application of log-rank test in comparing two survival curves as a greater percentage of censoring into the samples is introduced . To perform this study, real entomological data of males and females of *Ceraeochrysa cubana* , without censoring , were utilized and from them a sub-routine was set up in SAS® in which, by means of simulation, the random and type I censoring were introduced , by applying , thus, the log-rank test for comparing the resulting curves, this for different percentage of curese . By utilizing the real data , the test showed that for lifetimes without a censoring of males and females, the survival curves were different and as the type I censoring was introduced , they became similar from 30% of censoring. It was detected , through simulation which with an increase of the presence of censoring, the accuracy of the log-rank test has been decreasing. From the concepts of the central theorem of limit , it was possible to determine that , even with the introduction of censoring , the survival functions of males and females continued being different .In addition, the introduction of censoring to the data caused their survival curves to become different from the respective curves with the lifetimes whose information was complete . Besides the results shown above, this work still proposes a form of determining a correction factor of the estimators for samples with censored data, to approach them to the estimators for the samples without the presence of censoring.

¹ Adviser : Mário Javier Ferrua Vivanco – UFLA

1 INTRODUÇÃO

Estudos entomológicos, que avaliam dados de tempo de vida, são afetados em muitos casos pela perda de indivíduos por causas alheias ao experimento. A informação que corresponde aos indivíduos, perdidos geralmente não é utilizada na análise dos dados. Uma outra forma em que se perde informação é quando o estudo é encerrado num período preestabelecido de tempo. Nessas situações, os tempos de vida incompletos são chamados de censurados. No primeiro caso, ocorre a censura aleatória enquanto no segundo, a censura tipo I.

Em um experimento que analisa o tempo de vida de insetos, o descarte da informação de interesse pela perda de indivíduos é uma forma incorreta de analisar os dados. Isso ocorre porque a presença das observações incompletas no estudo se justifica, pelo fato de que elas fornecem, ainda, informações sobre o tempo de vida. Sem a presença destes dados, não seria possível realizar uma boa estimação de parâmetros e, assim, fazer uma análise adequada.

Para realizar uma análise correta e confiável de um experimento como o citado acima é fundamental a utilização das técnicas de Análise de Sobrevivência. Esta é a área da estatística que tem como principal interesse analisar dados de tempos de vida que envolvem algumas observações cuja informação é incompleta (dados censurados).

Estes dados são estudados de forma não-paramétrica, semi-paramétrica (modelos de regressão de Cox) e também com base em modelos estatísticos paramétricos.

Apesar da grande quantidade de pesquisas e trabalhos publicados sobre o tempo de vida na área da entomologia, não é comum encontrar nesses estudos, algum que tenha utilizado as técnicas de Análise de Sobrevida. Isto, provavelmente, ocorre devido à falta de conhecimento sobre este método, que passou a ser utilizado com frequência em diversas áreas, inclusive na agropecuária.

Além de mostrar as técnicas de Análise de Sobrevida, este trabalho tem também como objetivo estimar e comparar curvas de sobrevida para avaliar a possibilidade de aplicar mecanismos de censura, tanto a aleatória como a do tipo I, em dados entomológicos. Um outro objetivo é estudar a viabilidade da aplicação do teste log-rank na comparação de duas curvas de sobrevida, à medida que se introduz uma porcentagem cada vez maior de censura no experimento.

Para alcançar os objetivos foram utilizados dados reais de longevidade, de machos e fêmeas, da *Ceraeochrysa cubana*, acompanhada em ambiente laboratorial. Este inseto é um importante crisopídeo utilizado para o controle biológico natural de pragas de plantas cultivadas. Com base nesses dados reais, mediante um processo de simulação, foram geradas 5.000 amostras de tempos de vida para cada grupo. Utilizando-se os princípios do Teorema Central do Limite (TCL) foi feita uma análise comparativa para determinar se existem diferenças entre as funções de sobrevida obtidas com dados completos e as obtidas com a introdução de censura. Com as amostras geradas por simulação, foi possível comparar resultados obtidos pela análise que utiliza os conceitos do TCL, com resultados obtidos do teste log-rank, atualmente o teste estatístico mais utilizado para comparar duas funções de sobrevida. Além disso, foi possível também determinar empiricamente um fator de correção para os estimadores dos parâmetros relacionados com os dados censurados.

Na seção 2 será apresentada uma revisão bibliográfica dos assuntos relacionados com a metodologia da Análise de Sobrevivência e dos diversos estudos realizados com dados entomológicos. Nesta seção serão apresentados as principais definições e conceitos de Análise de Sobrevivência, mostrando as formas não-paramétrica e paramétrica para se estimar e comparar as funções de sobrevivência. O desenvolvimento metodológico utilizado neste estudo está na seção 3, em que se apresenta a forma de construção dos gráficos de linearização para a determinação do modelo probabilístico dos dados e os conceitos utilizados para desenvolver a sub-rotina de simulação de amostras para a longevidade de machos e fêmeas da *Ceraeochrysa cubana*, além de explicar os testes utilizados e o cálculo do fator de correção proposto. Os principais resultados são apresentados em tabelas e gráficos, chegando assim a conclusões claras e objetivas. No final, há uma seção de anexos, em que estão os dois tipos de sub-rotinas feitas em SAS® (Statistical Analysis System) para os casos da censura aleatória e a do tipo I. Sendo que cada um dos procedimentos realizados será destacado em forma de comentário.

2 REFERENCIAL TEÓRICO

A Análise de Sobrevivência é o conjunto de técnicas estatísticas para analisar tempos de vida que envolvem dados censurados (observações incompletas ou parciais sobre indivíduos que fizeram parte do experimento). Pois, a presença de censura inviabiliza a utilização de técnicas tradicionais da estatística, como Análise de Regressão e Planejamento de Experimentos.

Segundo Colosimo (2001), ao longo dos últimos vinte anos, a Análise de Sobrevivência tornou-se um dos métodos estatísticos mais utilizados, devido ao seu desenvolvimento e aprimoramento, combinado com computadores cada vez mais potentes.

Os dados de tempos de vida de indivíduos são aqueles medidos a partir do início do experimento até a perda ou falha destes. Nos estudos de dados tempo de vida, a variável resposta é o tempo até a ocorrência do evento de interesse, o chamado evento falha. Este tempo T é denominado **tempo de falha**, **tempo de vida** ou ainda **tempo de sobrevida**.

Segundo Soares & Colosimo (1995), a presença de dados censurados no estudo se justifica porque, mesmo sendo incompletos, eles nos fornecem informações sobre o tempo de vida e a omissão das censuras no cálculo das estatísticas de interesse nos levariam a conclusões viciadas.

A informação contida nestes dados é avaliada pela **função de sobrevivência**, denotada por $S(t)$, que é a probabilidade de um indivíduo não falhar até o tempo t e da **função de risco**, $h(t)$, que é a taxa instantânea de falha no tempo t de um indivíduo.

Segundo Lawless (1982), em termos probabilísticos, os tempos de sobrevivência se apresentam tanto em modelos contínuos como também em discretos.

a) Modelos contínuos

Seja T uma variável aleatória não negativa, representando o tempo de sobrevivência de um indivíduo e seja $f(t)$ a função densidade de probabilidade (f.d.p.) de T . A função de sobrevivência e a função de risco de T são expressas por:

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u) du$$

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Existem duas relações muito utilizadas em Análise de Sobrevivência entre $h(t)$ e $S(t)$. A partir da expressão de $h(t)$, temos que :

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

A segunda relação tem a ver com a função de risco acumulada, ou seja

$$H(t) = \int_0^t h(u) du = -\ln(S(t))$$

Segue-se então que

$$S(t) = e^{-H(t)}$$

Então nota-se que a f.d.p. de T pode ser representada pela função de risco $h(u)$,

$$f(t) = h(t) e^{-\int_0^t h(u) du}$$

Alisson (1995), demonstrou que tanto $f(t)$, $S(t)$ e $h(t)$, assim como, $F(t)$, (a função distribuição acumulada), representam, da mesma forma, a distribuição da variável aleatória T .

b) Modelos discretos

São modelos que correspondem a tempos de sobrevivência, em que os dados são agrupados, ou estão expressos em um número inteiro de unidades.

Assim, para a variável aleatória T , que assume os valores $t_1, t_2, \dots$, em que $0 \leq t_1 < t_2 < \dots$, sua função de sobrevivência é dada por:

$$S(t) = P[T \geq t] = \sum_{j: t_j \geq t} p(t_j)$$

Sua respectiva função de risco será dada da seguinte forma:

$$h(t_j) = P[T = t_j \mid T \geq t_j] = \frac{p(t_j)}{S(t_j)}, \quad j = 1, 2, \dots$$

Como $p(t_j) = S(t_j) - S(t_{j+1})$, temos então que

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)}, \quad j = 1, 2, \dots$$

Portanto, a função de sobrevivência pode ser representada como:

$$S(t) = \prod_{j: t_j \leq t} [1 - h(t_j)]$$

É ainda importante notar que, tanto para os casos de modelos contínuos como para os discretos, temos que $S(0) = 1$ e $S(\infty) = 0$.

Na prática, conforme afirmam Cox & Oakes (1984), modelos discretos são raramente encontrados. Usualmente arranjam-se os dados para uma distribuição contínua. Para muitos propósitos e, especificamente, para problemas de amostras que não envolvem covariáveis, têm-se como consequência pequenas incorreções nos dados que podem ser geralmente ignoradas.

Entre as diversas áreas de aplicação dos métodos de Análise de Sobrevivência, podem-se citar como exemplos:

- a área médica, em que eles são utilizados quando se quer identificar fatores de prognóstico ou de risco para uma determinada doença, bem como comparar tratamentos de interesse. Existem vários estudos sobre esse assunto na pesquisa clínica. Flehinger et al. (1998) mostram que a informação combinada entre os procedimentos utilizados para detectar o diagnóstico de falha e os procedimentos para diagnosticar as falhas nos casos mascarados providenciam estimativas sobre as funções de sobrevivência dos indivíduos em estudo. Teixeira et al.(2000) avaliaram a radioterapia e alguns fatores prognósticos associados à resposta terapêutica e à sobrevida em mulheres com carcinoma de colo uterino. Esses autores concluíram que a sobrevida total observada em cinco anos foi de 29%, um resultado que confirma o mau prognóstico para estes casos. Gamel et al. (2000) apresentaram um algoritmo que providencia estimativas de máxima verossimilhança para todos os parâmetros de um modelo de cura;
- a engenharia, na qual é conhecida como **Confiabilidade**, produtos, componentes e materiais são colocados em estudo para analisar características relacionadas à sua durabilidade. Valença (1994) utiliza a distribuição gama generalizada para discriminar modelos paramétricos. Freitas & Colosimo (1997) conceituam e definem a aplicação dos testes de vida acelerados. Carneiro (1999) apresenta uma metodologia para o mapeamento dos aspectos de confiabilidade por meio de dados de problemas de inspeção e ensaios de recebimento de equipamentos elétricos de alta tensão destinados às subestações do sistema elétrico da Companhia Paulista de Força e Luz, entre 1980 e 1995;
- as ciências sociais, em que se fazem estudos de nascimento e morte, duração de casamentos até o divórcio, entre outros. Elandt-Johnson & Johnson (1980) descrevem algumas aplicações nesta área.

Áreas menos exploradas estão relacionadas com estudos entomológicos, nas quais avaliam-se, entre outras, variáveis relacionadas com o tempo de sobrevivência de insetos quando o ambiente natural destes é, de alguma forma, manipulado. Diversos trabalhos desenvolvidos, que analisam variáveis do tipo mencionado anteriormente, não utilizam conceitos de Análise de Sobrevivência. Entre eles podem-se citar: Souza (1988) que, utilizando o teste de Duncan, mostrou que machos de *Nusalala uruguayana* alimentados com *T. citricidus*, *Dactynotus* sp., *Dactynotus* sp. + Sojinha^R, New Life^R e Novomilke^R, sobreviveram por um tempo significativamente maior em relação àqueles alimentados com Sojinha^R ou *B. brassicae*; Ribeiro (1988) apresentou a longevidade, ou seja, o tempo de vida desde o início da fase adulta até a sua morte, da *Chrysoperla externa* submetida a diversos tipos de dietas; Gonçalves (1990) fez um estudo comparando o ciclo total em dias de fêmeas e machos de *Podisus nigrolimbatus* e *Podisus connexivus*, alimentadas com lagartas de *Bombyx mori*, em que foi detectada uma diferença significativa entre as fêmeas dessas espécies; Silva (1991) afirma que a longevidade média dos crisopídeos é maior quando se utiliza uma dieta composta de uma fonte de proteína e outra de carboidrato; Nascimento (1996), também em estudo realizado em laboratório, encontrou que a longevidade média de *Podisus nigrispinus*, alimentados com lagartas de *Bombyx mori*, sadias foi de 39,8 dias, no caso das fêmeas e de 49,5 dias, no caso dos machos; entretanto, quando alimentados com lagartas de *Bombyx mori* infectadas com *Bacillus thuringiensis*, a longevidade passou para 38,3 dias, no caso das fêmeas e para 16,7 dias, no caso de machos.

Micheletti (1999), utilizando dados de longevidade retirados de levantamentos realizados em campo, estimou a sobrevivência de fêmeas e machos de *Cerconota anonella* ajustando a distribuição Weibull aos dados. Os estudos com dados de tempo de vida entomológicos, anteriormente mencionados, foram realizados utilizando-se a informação completa. Isto é, foi determinada e registrada a longevidade ou o ciclo total de vida para cada indivíduo. Nestes estudos, observações com informação incompleta são geralmente descartadas.

Já na Análise de Sobrevivência, a principal característica dos dados em estudo é a observação incompleta sobre o tempo de falha do indivíduo, ou seja, a **censura**. Allison (1995) menciona os três casos de censuras encontradas frequentemente: censura à direita, em que a ocorrência do evento de interesse acontece em um tempo superior ao parcialmente registrado, é o caso mais encontrado em estudos de tempo de vida; censura à esquerda, em que o evento acontece (ou aconteceu) em um tempo inferior ao tempo registrado e censura intervalar, que ocorre quando se tem o conhecimento que o tempo de falha ocorreu em um determinado intervalo de tempo. A censura intervalar é uma combinação das censuras à direita e a esquerda.

Segundo Colosimo (2001), este último tipo de censura ocorre em estudos agrônômicos, quando as visitas as unidades experimentais são especificadas entre datas distantes. Também quando o pesquisador visita diariamente o experimento, acarretando em uma unidade de medida, para o tempo de falha, grosseira para o estudo. Chalita (1997) apresenta um estudo completo para os dados com censura intervalar. Entre outros, a autora apresenta um experimento sobre a resistência de mangueiras.

Segundo Lawless (1982), existem três tipos de censura que são mais utilizadas e que são encontradas, tanto nos casos de censura à direita quanto de censura à esquerda. São elas:

- **Tipo I** – aquela onde estudo é encerrado após um período preestabelecido de tempo;
- **Tipo II** – aquela onde o estudo é encerrado após ter ocorrido a falha em um número preestabelecido de indivíduos;
- **Aleatória** – ocorre quando o indivíduo é retirado do estudo por causas alheias ao experimento.

Um tipo de censura, que não é exatamente como as mencionadas por Lawless (1982), porém semelhante com a Censura Tipo II, foi aplicada em estudo desenvolvido por Miramontes & De Souza (1996). Neste estudo foi analisada a sobrevivência de cupins colocados em tubos de ensaio hermeticamente fechados e reunidos em grupos de 1, 2, 3, 4, 8, 16 indivíduos. O experimento foi encerrado quando todos os cupins solitários morreram. Encontrou-se que a sobrevida foi significativamente maior para grupos maiores em relação a indivíduos isolados.

Nesse trabalho vamos nos fixar apenas na censura à direita, especificadamente na aleatória e na tipo I, provavelmente as que tem maior possibilidade de ser encontrada em estudos de tempo de vida entomológicos.

Para que haja uma análise confiável nos dados de tempo de vida com a presença de censura são destacadas três técnicas estatísticas: paramétrica, não-paramétrica e os modelos de regressão de Cox, também conhecida como técnica semi-paramétrica. A seguir, uma breve discussão sobre duas delas que serão utilizadas nesse trabalho: a não-paramétrica e a paramétrica.

2.1 Inferência não-paramétrica

Segundo Valença (1994), é comum sumarizar dados de vida em termos da função de sobrevivência empírica, que também representa um estimador não paramétrico da função de sobrevivência.

Caso não haja presença de dados censurados, a função de sobrevivência empírica é definida como:

$$\hat{S}(t) = \frac{o_i}{n_{o_i}}$$

sendo:

o_i o nº de indivíduos que não falharam até o tempo t_i ;

n_{o_i} o nº total de indivíduos no estudo.

Dois conhecidos estimadores não-paramétricos são a tabela de vida e o estimador produto limite, também conhecido como estimador de Kaplan-Meier, por estes dois autores terem sido os primeiros a discutirem suas propriedades. Colosimo (2001) afirma que a grande diferença entre os dois estimadores não-paramétricos de $S(t)$ está no número de intervalos utilizados para construção de cada um deles. O estimador de Kaplan-Meier é sempre baseado em um número de intervalos igual ao número de tempos de falhas distintos. Já a tabela de vida, os tempos de falha são agrupados em intervalos de forma arbitrária. Neste trabalho vamos apresentar somente o estimador de Kaplan-Meier que, conforme Valença (1994), é um estimador consistente sob condições gerais.

2.1.1 Estimador de Kaplan-Meier

Supondo que n indivíduos estão em um estudo, para os quais são registradas k falhas distintas, em que $k \leq n$, nos tempos $t_1 < t_2 < \dots < t_k$.

$S(t) = P(T > t)$ pode ser aproximada (ou estimada) por:

$$\hat{S}(t) = P(T \geq t_{j+1}) \text{ se } t \in [t_j; t_{j+1})$$

A $P(T \geq t_1)$, não $S(t_1)$ é igual a 1, se considerarmos que:

$$P(T \geq t_1) = \frac{o_1}{n_o} = 1$$

sendo:

o_1 o nº de indivíduos em risco em t_1 ;

n_o o nº total de indivíduos.

logo:

$$\hat{S}(t) = \frac{P(T \geq t_{j+1})}{P(T \geq t_1)}$$

Multiplicando e dividindo por $P(T \geq t_i), i = 2, 3, 4, \dots, j$, temos:

$$\hat{S}(t) = \frac{P(T \geq t_{j+1})}{P(T \geq t_1)} = \frac{P(T \geq t_2)}{P(T \geq t_1)} \cdot \frac{P(T \geq t_3)}{P(T \geq t_2)} \cdot \frac{P(T \geq t_4)}{P(T \geq t_3)} \dots \frac{P(T \geq t_{j+1})}{P(T \geq t_j)}$$

$$= P(T \geq t_2 \setminus T \geq t_1) \cdot P(T \geq t_3 \setminus T \geq t_2) \cdot P(T \geq t_4 \setminus T \geq t_3) \dots P(T \geq t_{j+1} \setminus T \geq t_j)$$

$$\Rightarrow \hat{S}(t) = \frac{n_{o_2}}{n_{o_1}} \cdot \frac{n_{o_3}}{n_{o_2}} \dots \frac{n_{o_{j+1}}}{n_{o_j}}$$

$$\Rightarrow \hat{S}(t) = \frac{n_{o_1} - n_{f_1}}{n_{o_1}} \cdot \frac{n_{o_2} - n_{f_2}}{n_{o_2}} \dots \frac{n_{o_j} - n_{f_j}}{n_{o_j}}$$

$$\Rightarrow \hat{S}(t) = \left(\frac{n_1 - d_1}{n_1} \right) \cdot \left(\frac{n_2 - d_2}{n_2} \right) \dots \left(\frac{n_j - d_j}{n_j} \right)$$

sendo:

n_{o_i} , o número de indivíduos em risco em t_i ;

n_{f_i} , o número de falhas em t_i .

Assim, a expressão geral do estimador de Kaplan-Meier pode ser apresentada, conforme Cox & Oakes (1984), utilizando-se a seguinte notação:

$$\hat{S}(t) = \prod_{\forall t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{\forall t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

em que:

d_j : nº de indivíduos que falharam no tempo t_j ;

n_j : nº de indivíduos sob risco (não falhou e não foi censurado) até o tempo t_j (exclusive);

$\frac{d_j}{n_j}$: indica a proporção de falhas em t_j ;

$1 - \frac{d_j}{n_j}$: indica a proporção de sobrevivência a t_j , depois de ter sobrevivido a $(j-1)$ tempos.

Kaplan & Meier (1958) demostraram que a expressão apresentada pode ser considerada como um estimador não-paramétrico de máxima verossimilhança para $S(t)$. Para isto, o autor construiu a função de verossimilhança observada, dada por:

$$L = \prod_{j=1}^k \left[\left(\prod_{i=1}^{\lambda_j} S(L_i^j) \right) \left[S(t_j) - S(t_j + 0) \right]^{d_j} \right] \prod_{i=1}^{\lambda_{k+1}} S(L_i^{k+1})$$

em que:

λ_j é o número de tempos L_i^j censurados no intervalo $[t_{j-1} ; t_j)$.

Devido a isso, esse estimador passou a ser o mais utilizado em estudos de Análise de Sobrevivência.

Técnicas não-paramétricas são utilizadas quando existem dificuldades na identificação de um modelo paramétrico para os dados em estudos. Mas, conforme Bolfarine et al. (1991), as técnicas paramétricas são em geral bem mais eficientes que as técnicas não-paramétricas. Além disso na maioria dos casos, os dados possuem informação suficiente para a identificação de um modelo que seja razoavelmente adequado.

2.2 Inferência paramétrica

Os modelos paramétricos são aqueles em que a análise dos dados requer a especificação de uma distribuição de probabilidade para o tempo de falha. Uma vez escolhida a distribuição de probabilidade que melhor descreve o comportamento da curva de sobrevivência é então possível estimar respectivamente a função de sobrevivência e a função de risco. Na literatura são encontrados alguns modelos que são usados freqüentemente. Lawless (1982), Bolfarine et al. (1991) e Valença (1994) apresentam, entre outros, os seguintes: exponencial, Weibull, gama, log-normal e gama generalizada. A seguir, apresentaremos detalhes sobre estes modelos paramétricos contínuos.

2.2.1 Distribuição exponencial

Uma variável aleatória T com distribuição exponencial com parâmetro λ tem f.d.p. dada por:

$$f(t) = \lambda e^{-\lambda t}, \quad \lambda > 0, t > 0$$

Conseqüentemente, temos que sua função de sobrevivência é:

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

Por ser uma distribuição que possui estrutura simples, já que possui um único parâmetro, ela é usada frequentemente, tanto na área médica como na industrial.

Uma característica importante desta distribuição é o fato de possuir função de risco constante, ou seja:

$$h(t) = \lambda$$

A exponencial é a única que possui essa característica. Isso significa, segundo Valença (1994), que, para este modelo, a chance condicional de falha em um intervalo de tempo de comprimento estipulado é a mesma, independentemente do tempo em que o item venha sendo testado. Assim, o envelhecimento do item não estaria influenciando no risco de falha instantânea.

Fazendo $Y = \ln(T)$, tem-se a f.d.p. de Y dada por:

$$f_Y(y) = \exp(y - \alpha - e^{y-\alpha}), \quad y \in \mathbb{R}$$

em que $\alpha = \ln \lambda$. Por ser α um parâmetro de locação, pode-se representar Y da seguinte forma:

$$Y = \alpha + W$$

sendo W uma variável aleatória com distribuição valor extremo padrão, com suas respectivas f.d.p. e função de sobrevivência:

$$f_W(w) = \exp(w - e^w)$$

$$S_W(w) = \exp(-e^w), \quad w \in \mathbb{R}$$

2.2.2 Distribuição de Weibull

Esta distribuição foi proposta por Weibull (1951) em estudos sobre o tempo de falha devido à fadiga de metais.

Lawless (1982) o descreve como o modelo paramétrico mais utilizado no ajuste de tempo de vida, na área médica e na engenharia. Colosimo (2001) afirma que a popularidade desta distribuição deve-se ao fato dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de risco é monótona. Ou seja, ela é crescente, decrescente ou constante.

A f.d.p. da Weibull é dada por:

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} e^{-\left(t/\alpha\right)^\beta}, t \geq 0$$

Os parâmetros de forma β e o de escala α são positivos.

O parâmetro de escala tem a mesma unidade de t e o de forma não tem unidade.

Sua função de sobrevivência é a seguinte:

$$S(t) = e^{-\left(t/\alpha\right)^\beta}, t > 0$$

A função de risco de T é dada por:

$$h(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1}$$

Portanto, é fácil perceber que a distribuição exponencial é um caso particular da distribuição Weibull, quando $\beta = 1$.

Fazendo $Y = \ln(T)$, tem-se que Y segue uma distribuição valor extremo com f.d.p. e função de sobrevivência dadas abaixo:

$$f_Y(y) = \frac{1}{\sigma} \exp \left[\frac{y - \delta}{\sigma} - \exp \left(\frac{y - \delta}{\sigma} \right) \right], \quad y \in \mathbb{R}$$

$$S_Y(y) = \exp \left(-\exp \left(\frac{y - \delta}{\sigma} \right) \right)$$

sendo $\sigma = \frac{1}{\beta}$ e $\delta = \log \alpha$ (β e α parâmetros da Weibull). Como a distribuição valor extremo pertence à família de modelos de locação e escala, pode-se, então, escrever Y como:

$$Y = \delta + \sigma W$$

em que W tem distribuição valor extremo padrão, que é um caso particular da distribuição valor extremo, quando $\delta = 0$ e $\sigma = 1$.

2.2.3 Distribuição log-normal

Se T é uma variável aleatória, que tem distribuição log-normal com parâmetros μ e σ^2 , sendo $-\infty < \mu < \infty$ e $\sigma^2 > 0$, então sua f.d.p. é expressa da seguinte forma:

$$f(t) = \frac{1}{(2\pi)^{1/2} \sigma t} e^{\left\{-1/2 \left(\frac{\ln(t) - \mu}{\sigma}\right)^2\right\}}, t > 0$$

Sua principal característica é o fato de $Y = \ln(T)$ possuir distribuição normal com média μ e variância σ^2 . Assim, tem-se que $Y = \mu + \sigma Z$, em que Z segue uma distribuição normal padronizada, com função de distribuição dada por:

$$\Phi(z) = \int_{-\infty}^z \frac{1}{(2\pi)^{1/2}} e^{-u^2/2} du, \quad z \in \mathbb{R}$$

Sua função de sobrevivência fica, então, relacionada com a $\Phi(z)$, segundo:

$$S(t) = 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$$

Conseqüentemente, a função de risco também não apresenta uma forma fechada. Pode ser demonstrado, que quando $t = 0$, então $h(t) = 0$, crescendo a continuação para um máximo e, posteriormente, decrescendo até aproximar-se de 0 quando $t \rightarrow \infty$.

A distribuição log-normal possui uma estreita relação com a distribuição normal, já que o logaritmo de uma variável com distribuição log-normal, com parâmetros μ e σ , tem uma distribuição normal com média μ e desvio padrão σ . Isto significa que dados que seguem uma distribuição log-normal podem ser analisados segundo uma distribuição normal, se utilizarmos o logaritmo dos dados no lugar dos dados originais.

2.2.4 Distribuição gama

Uma variável aleatória T , com distribuição gama, tem sua f.d.p dada por:

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

em que $\alpha > 0$ e $k > 0$, sendo α o parâmetro de escala e k o parâmetro de forma.

Sua função de sobrevivência é expressada como:

$$S(t) = 1 - I(k, t/\alpha),$$

em que $I(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u} du$ é a função gama incompleta.

A função de risco é dada por:

$$h(t) = \left[\frac{1}{\Gamma(k)} \right] \left(\frac{t^{k-1}}{\alpha^k} \right) \frac{e^{(-t/\alpha)}}{[1 - I(K, t/\alpha)]}$$

sendo que:

$h(t)$ é monótona crescente para $k > 1$,

$h(0)$ é constante, com $h(0) = 0$ e para $k = 1$,

$h(t)$ é monótona decrescente para $0 < k < 1$.

Duas situações especiais ocorrem nessa distribuição:

Se $\alpha = 1$, então T tem distribuição gama com apenas um parâmetro, isto é:

$$f(t) = \frac{1}{\Gamma(k)} t^{k-1} e^{-t}, \quad t > 0$$

Se $k = 1$, a distribuição gama se reduz à distribuição exponencial.

Entre as distribuições gama e a exponencial há uma importante relação, que é o fato de a soma de exponenciais independentes e identicamente distribuídas possuir distribuição gama.

2.2.5 Distribuição gama generalizada

Uma variável aleatória T com distribuição gama generalizada tem f.d.p. dada por:

$$f(t) = \frac{\beta}{\alpha \Gamma(k)} (t/\alpha)^{\beta k - 1} e^{-(t/\alpha)^\beta}, \quad t > 0$$

em que $\alpha, \beta, k > 0$

Sua característica principal é o fato de representar uma família paramétrica que possui como caso particulares outras distribuições intensamente utilizadas em Análise de Sobrevida.

As funções de sobrevivência e de risco são expressas, respectivamente, da seguinte forma:

$$S(t) = 1 - I[k, (t/\alpha)^\beta]$$

$$h(t) = \frac{t^{\beta k - 1} e^{-(t/\alpha)^\beta}}{\int_0^\infty x^{\beta k - 1} e^{-(x/\alpha)^\beta} dx}$$

Convém aqui destacar que as distribuições apresentadas em 2.2.1, 2.2.2, 2.2.3 e 2.2.4 são casos particulares da gama generalizada, como é detalhado a seguir:

- i) se $\beta = k = 1$, então T tem distribuição exponencial;
- ii) se $k = 1$, T tem distribuição Weibull;
- iii) se $\beta = 1$, T tem distribuição gama;
- iv) se $k \rightarrow \infty$, T tem distribuição log-normal.

Valença (1994) apresenta uma discussão mais detalhada sobre esta distribuição.

Utilizando inferência não-paramétrica ou paramétrica, o objetivo principal dos estudos de tempo de vida é fazer a comparação entre as curvas de sobrevivência e verificar se há covariáveis que influenciam nas diferenças, caso existam, entre essas curvas.

2.3 Comparação de duas curvas de sobrevivência

É importante mencionar que, na literatura, o mais comum é trabalhar com testes baseados nos postos de observações, os chamados testes não-paramétricos ou assintoticamente não-paramétricos, para comparar duas curvas de sobrevivência. Um ponto favorável à aplicação dos testes não-paramétricos, segundo Oliveira (1981), é que a região de rejeição é a mesma para qualquer que seja a distribuição real das observações. Testes paramétricos para modelos exponenciais também são frequentemente encontrados. A seguir, serão destacados alguns testes utilizados para comparar duas curvas de sobrevivência, encontrados em Chalita (1992) e Colosimo (2001).

2.3.1 Teste de Gehan

Gehan (1965) propôs uma generalização do teste de Wilcoxon para comparar a função de sobrevivência de dois grupos. O autor supõe que as hipótese a ser testada é: $H_0 : F_1(t) = F_2(t)$ versus $H_1 : F_1(t) < F_2(t)$, sendo $F_1(t)$ e $F_2(t)$ as funções distribuição acumuladas (discretas ou contínuas) dos dois grupos em estudo. Este teste mostrou-se consistente e sua estatística tem distribuição assintoticamente normal com média zero. Quando H_0 é verdadeira, o teste não é seriamente afetado por diferenças reais na porcentagem de censura nos dois grupos. Define-se U_{ij} , como sendo:

$$U_{ij} = \begin{cases} -1, & \text{se } x_i < y_j \text{ ou } x_i \leq y'_j \\ 1, & \text{se } x_i > y_j \text{ ou } x'_i \geq y_j \\ 0, & \text{c. c.} \end{cases}$$

sendo que, x_i e y_j são os tempos de falhas e x'_i e y'_j , são os tempos para a censura. A estatística do teste é:

$$W = \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} U_{ij}$$

Chalita (1992) descreve este teste mais detalhadamente. Breslow (1970) estendeu esse estudo para mais de dois grupos.

2.3.2 Teste F de Cox

Cox (1953) sugere um teste F que pode ser empregado para testar diferentes tratamentos com presença ou ausência de censura, quando pode ser assumido que os tempos de falha seguem uma distribuição exponencial em ambos os grupos. Chalita (1992) apresenta o teste da seguinte forma:

Suponha que a hipótese a ser testada seja

$H_0 : \lambda_1 = \lambda_2$ versus

$H_1 : \lambda_1 < \lambda_2$, $H_1 : \lambda_1 > \lambda_2$, $H_1 : \lambda_1 \neq \lambda_2$

Então, supondo que tenham n_1 e n_2 indivíduos nos grupos 1 e 2, respectivamente, $X_1, \dots, X_{r_1}$ não censurados e $X'_{r_1+1}, \dots, X'_{n_1}$ censurados no grupo 1 e $Y_1, \dots, Y_{r_2}$ não censurados e $Y'_{r_2+1}, \dots, Y'_{n_2}$ censurados no grupo 2, onde

$$\bar{t}_1 = \frac{\sum_{i=1}^{r_1} x_i + \sum_{i=r_1+1}^{n_1} x'_i}{r_1} \text{ e } \bar{t}_2 = \frac{\sum_{i=1}^{r_2} y_i + \sum_{i=r_2+1}^{n_2} y'_i}{r_2}$$

Um teste eficiente para as hipóteses propostas é tomar $\bar{t}_1 / \bar{t}_2$ como tendo uma distribuição F com $(2r_1, 2r_2)$ graus de liberdade.

As regras de decisão para o teste são as seguintes:

- a) unilateral à esquerda, rejeita-se H_0 se $\bar{t}_1 / \bar{t}_2 < F_{(2r_1, 2r_2, 1-\alpha)}$;
- b) unilateral à direita, rejeita-se H_0 se $\bar{t}_1 / \bar{t}_2 > F_{(2r_1, 2r_2, \alpha)}$;
- c) bilateral, rejeita-se H_0 se $\bar{t}_1 / \bar{t}_2 < F_{(2r_1, 2r_2, 1-\frac{\alpha}{2})}$ ou se $\bar{t}_1 / \bar{t}_2 > F_{(2r_1, 2r_2, \frac{\alpha}{2})}$,

em que α é o nível da significância.

Lee (1980), citado por Chalita (1992), mostra que se o tempo de sobrevivência T tem distribuição Weibull, com parâmetro de forma β , então T^β terá uma distribuição exponencial. Assim, quando as observações em cada amostra são elevadas à β (sempre que β seja conhecido) pode-se utilizar o teste F de Cox para comparar as duas amostras

Um outro teste, que é atualmente o teste mais utilizado em Análise de Sobrevivência, é o teste log-rank. Ele é apropriado quando a razão das funções de risco a serem comparadas é aproximadamente constante; isto significa que as populações têm propriedade de riscos proporcionais. Este teste será apropriadamente discutido na seção 3 (Material e Métodos), em que será utilizado para análise dos dados propostos nesse trabalho.

Até agora foi realizada uma abordagem sobre populações homogêneas, mas, na prática, há muitas situações em que os dados de tempo de vida podem ser afetados por variáveis regressoras, que também são chamadas de variáveis explicativas ou covariáveis. Nesse caso, podem ser ajustados modelos de regressão. A seguir, serão apresentados brevemente os modelos de regressão mais utilizados. É importante mencionar que, neste trabalho, não foram usados modelos de regressão, já que a análise foi feita com dados obtidos em laboratório. Isto é, trabalhou-se com indivíduos homogêneos e todos os fatores que poderiam afetar a variável em estudo foram controlados.

2.4 Modelos de regressão

Considere uma variável aleatória T , tempo de falha de indivíduo e seja $\mathbf{x}$ um vetor formado por observações de $p-1$ covariáveis $(x_1, \dots, x_{p-1})$, que podem ser quantitativas ou qualitativas, e $x_0 = 1$. Um modelo de regressão é uma maneira de determinar o relacionamento entre T e $\mathbf{x}$.

Na literatura, são encontradas duas importantes classes de modelos de regressão:

- riscos proporcionais para T ;
- locação e escala para $\ln(T)$.

Serão abordados aqui os modelos de regressão de locação e escala.

2.4.1 Modelos de locação e escala

Segundo Valença (1994), esta classe de modelos de regressão se caracteriza pelo fato de $Y = \ln(T)$ ter uma distribuição com parâmetros de locação $\mu(\mathbf{x})$ dependendo das covariáveis, e um parâmetro de escala σ , constante. Assim, pode ser escrita como:

$$Y = \mu(\mathbf{x}) + \sigma W$$

em que $\sigma > 0$ e W tem distribuição que não depende de $\mathbf{x}$.

A função de sobrevivência de Y dado $\mathbf{x}$ é da forma $G\left(\frac{y - \mu(\mathbf{x})}{\sigma}\right)$, em que $G(\cdot)$ é a função de sobrevivência de W , sendo expressa da seguinte forma:

$$S(t \setminus \mathbf{x}) = G\left[\log(t / \alpha(\mathbf{x}))^\delta\right]$$

em que $\alpha(\mathbf{x}) = e^{\mu(\mathbf{x})}$ e $\delta = \frac{1}{\sigma}$.

Uma característica conhecida desse modelo é que as covariáveis atuam multiplicativamente sobre T . Dependendo do conjunto de dados em estudo, pode-se assumir várias formas para $\mu(\mathbf{x})$. A maneira natural é assumir $\mu(\mathbf{x}) = \mathbf{x}\beta$, em que $\beta = (\beta_0, \dots, \beta_{p-1})'$ é um vetor de parâmetros desconhecidos. Utilizando esta suposição, o modelo de locação e escala é um modelo log-linear para T com resíduo W .

A seguir, serão apresentados os modelos log-linear que são extensões das distribuições exponencial, Weibull e log-normal, citados em Valença (1994).

2.4.1.1 Modelo de regressão exponencial

Considere T dado $\mathbf{x}$ com distribuição exponencial e seja $Y = \ln(T)$. O modelo de regressão log-linear para T é dado por :

$$Y = \mathbf{x}\beta + W,$$

em que W tem distribuição valor extremo padrão, sendo que a f.d.p. e a função de sobrevivência de Y dado $\mathbf{x}$ são, respectivamente :

$$f_Y(y \setminus \mathbf{x}) = \exp[y - \mathbf{x}\beta - \exp(y - \mathbf{x}\beta)], \quad y \in \mathbb{R};$$

e

$$S_Y(y \setminus \mathbf{x}) = \exp[-\exp(y - \mathbf{x}\beta)]$$

2.4.1.2 Modelo de regressão Weibull

Se T dado $\mathbf{x}$ tem distribuição Weibull com parâmetros $\alpha(\mathbf{x}) = e^{\mathbf{x}\beta}$ e $b = 1/\sigma$, o modelo de regressão log-linear é da forma:

$$Y = \mathbf{x}\beta + \sigma W,$$

em que W tem distribuição valor extremo padrão. A f.d.p. e a função de sobrevivência para Y dado $\mathbf{x}$ são, respectivamente:

$$f_Y(y \setminus \mathbf{x}) = \frac{1}{\sigma} \exp \left[\left(\frac{y - \mathbf{x}\beta}{\sigma} \right) - \exp \left(\frac{y - \mathbf{x}\beta}{\sigma} \right) \right], \quad y \in \mathfrak{R},$$

$$S_Y(y \setminus \mathbf{x}) = \exp \left\{ -\exp \left(\frac{y - \mathbf{x}\beta}{\sigma} \right) \right\}$$

2.4.1.3 Modelo de regressão log-normal

O modelo de regressão log-linear, cujo tempo de falha T dado $\mathbf{x}$ tem distribuição log-normal, pode ser representado por:

$$Y = \mathbf{x}\beta + \sigma Z$$

em que Z tem distribuição normal padrão. A densidade e a função de sobrevivência para este modelo são:

$$f_Y(y \setminus \mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{y - \mathbf{x}\beta}{\sigma} \right)^2 \right], \quad y \in \mathfrak{R},$$

$$S_Y(y \setminus \mathbf{x}) = 1 - \Phi \left(\frac{y - \mathbf{x}\beta}{\sigma} \right)$$

Em Allison (1995), é encontrado um resumo das distribuições de W para as principais distribuições do modelo de locação e escala.

Antes de encerrar esta seção, é bom esclarecer que muitos dos experimentos entomológicos são realizados em laboratórios, em que covariáveis que podem influenciar na ocorrência do evento de interesse são controladas. Um exemplo desse tipo é o de Silva (1991), que determina a longevidade em dias de um crisopídeo dentro de laboratório com temperatura e umidade relativa do ar controlada.

Em estudos agropecuários ainda não existe no Brasil, um grande número de trabalhos realizados utilizando técnicas de Análise de Sobrevivência. As principais referências para este estudo são os trabalhos de Chalita (1992) e Chalita (1997). No primeiro, simulou-se dois conjuntos de dados para exemplificar a utilização da inferência paramétrica e não-paramétrica em dados de tempo de vida, onde também foi calculada a eficiência assintótica do modelo paramétrico em relação ao não-paramétrico. Já no segundo trabalho, a autora faz um estudo sobre os modelos mais apropriados para dados agrupados e, ao mesmo, tempo censurados. Ou seja, dados em que se observa em que intervalo ocorreu a falha, acarretando, assim, um grande número de empates.

Na Universidade Federal de Lavras (UFLA), apesar da grande quantidade de pesquisas e trabalhos publicados sobre tempo de vida de insetos, nenhum deles aplicou esta técnica. Todavia, ela é extremamente prática e útil para a análise de dados em que o tempo de vida do indivíduo é o objeto de interesse. Este trabalho trata da aplicação das técnicas de Análise de Sobrevivência, introduzindo mecanismos de censura, em um estudo realizado na UFLA, em que foram medidos a longevidade dos insetos. Esse experimento apresenta a informação completa sobre o tempo de falha dos indivíduos.

3 MATERIAL E MÉTODOS

Visando à comparação de curvas de sobrevivência, por meio de um conjunto de dados de tempo de vida entomológicos retirados de um estudo realizado por Silva (1991), em que foi determinada a longevidade em dias da *Ceraeochrysa cubana*, determinou-se a distribuição que se ajusta melhor aos dados. Para isso, foram utilizados gráficos de linearização da função de sobrevivência.

Determinada a distribuição dos dados e utilizando uma sub-rotina desenvolvida em SAS® por Vivanco (1994), foi possível estimar os parâmetros da distribuição dos tempos de falha, pelo método da máxima verossimilhança, utilizando o método de Newton-Raphson. Com os estimadores encontrados geraram-se, pelo método de simulação de Monte Carlo, 5.000 amostras de tamanho igual ao das amostras originais, ou seja, 40 e 31 indivíduos machos e fêmeas, respectivamente. Posteriormente, nas amostras simuladas, foram introduzidas porcentagens de censura, aleatória e tipo I, iguais a 10%, 20%, 30%, 40%, 50%.

Para todas as amostras geradas, censuradas e não censuradas foram determinadas as estimativas de máxima verossimilhança. A partir dos fundamentos do teorema central do limite (TCL), foi possível construir um intervalo de confiança para a diferença das médias dos estimadores correspondentes às amostras censuradas e não censuradas. Com os mesmos fundamentos do TCL foram testadas hipóteses de comparação de médias de estimadores para machos *versus* fêmeas, primeiro utilizando as amostras não censuradas e depois as censuradas.

Os resultados do teste de hipótese de comparação de médias de estimadores de machos e fêmeas, para amostras censuradas, serviram como referência para avaliar a eficiência do teste log-rank (o mais utilizado na comparação de curvas de sobrevivência).

Por outro lado, em muitos casos, a aplicação de censura pode criar um vício muito grande (subestimação ou superestimação) na estimação dos parâmetros do modelo proposto para os dados de tempo de vida. Assim, foi proposta uma alternativa empírica para corrigir as estimativas obtidas com dados censurados e, dessa forma, aproximá-las das estimativas provavelmente obtidas com os dados sem censura. Tal alternativa de correção foi denominada "fator de correção para estimativas censuradas".

3.1 Materiais

Para alcançar os objetivos propostos nesse trabalho, foram utilizados dados reais de longevidade. Ou seja, são dados do início da fase adulta até a sua morte, de machos e fêmeas, da *Ceraeochrysa cubana*, acompanhada em ambiente laboratorial, em um estudo realizado por Silva (1991) (Tabela 1). Este inseto é um importante crisopídeo utilizado para o controle biológico natural de pragas de plantas cultivadas

TABELA 1 Longevidade em dias da *Ceraeochrysa cubana* sob temperatura $25 \pm 2^{\circ}\text{C}$, UR $70 \pm 10\%$, fotofase de 12 horas.

N ^o	MACHO	FÊMEA
1	100	111
2	165	58
3	77	87
4	94	117
5	88	39
6	142	84
7	200	87
8	100	106
9	122	56
10	67	131
11	59	34
12	200	94
13	142	118
14	195	123
15	42	68
16	63	75
17	39	103
18	156	63
19	117	32
20	47	41
21	42	94
22	76	78
23	49	86
24	205	44
25	163	109
26	199	50
27	67	138
28	44	33
29	110	129
30	110	150
31	121	70
32	184	
33	132	
34	188	
35	190	
36	82	
37	83	
38	115	
39	139	
40	201	

3.2 Determinação do modelo paramétrico adequado

Quando escolhe-se a inferência paramétrica para analisar dados de tempo de vida, a escolha do modelo passa a ser uma etapa importante na análise.

A seguir será apresentado um método que é intensamente utilizado para a discriminação de modelos.

3.2.1 Gráficos de linearização da função de sobrevivência

Consiste em construir gráficos que sejam aproximadamente lineares, se o modelo proposto for apropriado. Para a construção do gráfico utiliza-se uma transformação que lineariza o modelo proposto. Conseqüentemente, se o modelo for adequado, os dados amostrais aproximam-se de uma reta como resultado final.

A seguir apresentam-se as formas para construção dos gráficos de linearização de três modelos paramétricos, encontradas em Freitas e Colosimo (1997): exponencial, Weibull e log-normal.

a) Distribuição Weibull: $\ln[-\ln(\hat{S}(t))]$ versus $\ln(t)$

b) Distribuição exponencial: $-\ln(\hat{S}(t))$ versus t

c) Distribuição log-normal: $\Phi^{-1}(\hat{S}(t))$ versus $\ln(t)$

em que, $\hat{S}(t)$ é o estimador de Kaplan-Meier para a função de sobrevivência e $\Phi(.)$ é a função de densidade da distribuição normal padrão.

3.3 Máxima verossimilhança

Uma estimativa pontual confiável, conforme Spiegel (1978), é obtida por meio de uma técnica conhecida como Estimação de Máxima Verossimilhança.

Suponha que a população tenha função densidade que contenha um parâmetro populacional θ , a ser estimado por meio de determinada estatística. A função de densidade então, pode ser denotada por $f(x, \theta)$. Admitindo uma amostra aleatória de tamanho n : $x_1, x_2, \dots, x_n$, a função de densidade conjunta dessa amostra será:

$$L = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

que é chamada Função de Verossimilhança. O estimador de máxima verossimilhança pode ser obtido maximizando o logaritmo da função de verossimilhança. Uma forma de maximizar e achar os pontos críticos é derivando em relação a θ e igualando-a a zero. Assim, temos que:

$$\frac{1}{f(x_1, \theta)} \frac{\partial f(x_1, \theta)}{\partial \theta} + \dots + \frac{1}{f(x_n, \theta)} \frac{\partial f(x_n, \theta)}{\partial \theta} = 0$$

Desta equação, pode-se obter θ em termos da amostra aleatória.

O método é possível de generalização. No caso de vários parâmetros, são feitas as derivadas parciais em relação a cada um deles e, posteriormente, iguala-se a zero.

Na maioria das vezes, o cálculo para a obtenção dos estimadores de máxima verossimilhança não é possível analiticamente. Então, métodos iterativos de aproximação numérica, como o método de Newton-Raphson, são utilizados.

3.3.1 Máxima verossimilhança com a presença de censura

Nesta seção serão apresentados os casos de estimação para distribuições contínuas, segundo Cox e Oakes (1984).

Quando se trabalha com censura não informativa, isto é, quando pode-se assumir que o tempo até a ocorrência do evento em estudo é o mesmo para todos os indivíduos, ocorrendo ou não censura, a função de verossimilhança será a mesma para censura tipo I, tipo II e aleatória e é determinada da seguinte forma:

Um indivíduo observado até o tempo de falha t contribui com o termo $f(t; \theta)$ para a verossimilhança, ou seja, a densidade de falha em t , sendo θ um vetor de parâmetros associados à densidade da variável aleatória T ($\theta = (\theta_1, \theta_2, \dots, \theta_n)$). A contribuição do indivíduo, quando o tempo de sobrevivência é censurado no tempo c , é a probabilidade de sobrevivência até c , $S(c)$. Sendo assim, a função de verossimilhança será dada por:

$$L(\theta) = \prod_{i \in F} f(t_i; \theta) \prod_{i \in c} S(t_i; \theta)$$

em que, "F" representa falha e "c" censura.

O processo de maximização se torna mais simples quando trabalha-se com o logaritmo natural de $L(\theta)$, chamada de Função Log-verossimilhança, dada por:

$$l = \sum_{i \in F} \ln f(t_i; \theta) + \sum_{i \in c} \ln S(t_i; \theta) \quad (a)$$

Em termos de falhas ou tempos censurados observados, fazendo $x_i = \min(t_i; c_i)$, a função log-verossimilhança pode ser escrita como:

$$l = \sum_{i \in F} \ln f(x_i; \theta) + \sum_{i \in c} \ln S(x_i; \theta)$$

Sabendo que $f(t) = h(t)S(t)$, l pode ser reescrita como:

$$l = \sum_{i \in F} \ln [h(x_i, \theta)S(x_i, \theta)] + \sum_{i \in c} \ln S(x_i, \theta)$$

portanto,

$$l = \sum_{i \in F} \ln h(x_i; \theta) + \sum_{i=1}^n \ln S(x_i, \theta) \quad (b)$$

A determinação dos estimadores de máxima verossimilhança requer a maximização das funções log-verossimilhança (a) ou (b), ou seja, requer a solução das equações:

$$\frac{\partial l(\theta)}{\partial \theta_j} = 0$$

Como geralmente os estimadores não são encontrados analiticamente, neste caso há também a necessidade da utilização de métodos iterativos como o de Newton-Raphson. Em Lawless (1982), onde são encontradas as funções de verossimilhança para os principais modelos contínuos de tempo de vida com presença de censura, o autor faz também uma descrição sobre o método de Newton-Raphson.

No caso da distribuição Weibull, os parâmetros de forma e escala serão estimados segundo o método de máxima verossimilhança da seguinte maneira:

$$\begin{aligned}\ln L(\alpha, \beta) &= \sum_{i \in F} \ln \left[\frac{\beta}{\alpha} \left(\frac{t_i}{\alpha} \right)^{\beta-1} \right] + \sum_{i=1}^n \ln e^{\left[-\left(\frac{t_i}{\alpha} \right)^\beta \right]} \\ \Rightarrow \ln L(\alpha, \beta) &= \sum_{i \in F} \{ \ln \beta - \ln \alpha + (\beta-1) [\ln t_i - \ln \alpha] \} + \sum_{i=1}^n - \left(\frac{t_i}{\alpha} \right)^\beta \\ \Rightarrow \ln L(\alpha, \beta) &= r \ln \beta - r \ln \alpha + (\beta-1) \sum_{i \in F} \ln t_i - \sum_{i=1}^n \left(\frac{t_i}{\alpha} \right)^\beta\end{aligned}$$

em que r é o número de falhas.

Pelo processo de maximização pode-se chegar aos seguintes resultados:

$$\hat{\beta} = \frac{r}{r \ln \alpha - \sum_{i \in F} \ln t_i + \sum_{i=1}^n \left(\frac{t_i}{\alpha} \right)^\beta \ln \left(\frac{t_i}{\alpha} \right)} \quad \text{e} \quad \hat{\alpha} = \left(\frac{1}{r} \sum_{i=1}^n t_i^{\hat{\beta}} \right)^{\frac{1}{\hat{\beta}}}$$

Pode-se observar que os estimadores obtidos após o processo de estimação pontual não possuem uma expressão fechada. A obtenção de estimadores pela aproximação numérica, segundo o método de Newton-Raphson, pode ser encontrada na seção de Anexos (Simulação 1 e Simulação 2) em que foi utilizado o procedimento PROC LIFEREG do SAS®.

3.4 Simulação das amostras

Tornou-se cada vez mais freqüente em estudos científicos o uso dos métodos de simulação para estudar novos procedimentos estatísticos ou para comparar resultados de diferentes técnicas estatísticas já existentes. Os processos de simulação que envolvem componentes aleatórios são pertencentes ao método de Monte Carlo. Santos (2001) afirma que o método de Monte Carlo, de uma maneira bastante simplificada, é um algoritmo que consiste em simular dados, partindo de uma seqüência pseudo-aleatória, baseada na distribuição uniforme (0,1). Este método é aplicado no presente estudo, considerando o teorema da transformação integral da probabilidade, que é demonstrado conforme Mood et al. (1974), da seguinte forma:

Teorema 1: Teorema da transformação integral da probabilidade

1ª) Se a variável aleatória contínua $X \sim F_X(\cdot)$, então:

$$U = F_X(X) \sim U(0,1);$$

2ª) Se $U \sim U(0,1)$ e $X = F_X^{-1}(U)$, então $X \sim F_X(\cdot)$.

Demonstração:

1ª) Sendo X uma variável aleatória com distribuição $F_X(\cdot)$ e $U = F_X(X)$, então qual a distribuição de U ?

$$F_U(u) = P(U \leq u) = P(F_X(X) \leq u) = P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u.$$

Logo,

$$U \sim U(0,1).$$

2^o) Sendo $U \sim U(0,1)$ e $X = F_X^{-1}(U)$, então qual distribuição de X ?

$$P(X \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x).$$

Logo, $F_X(\cdot)$ é a função de distribuição acumulada de X .

O Teorema 1, associado ao método de Monte Carlo, permitiu simular 5.000 amostras de tempos de vida (longevidade) de machos e fêmeas, de tamanho 40 e 31, respectivamente, de acordo com o modelo probabilístico determinado segundo a metodologia explicada em 3.2.1.

3.5 Intervalos de confiança e testes de hipóteses para comparação de médias de estimadores

Com as 5.000 amostras geradas, pode-se afirmar, pelo teorema central do limite, que a média dos 5.000 estimadores obtidos para cada parâmetro correspondente à distribuição de tempos de vida, sem censura e censurados, segue uma distribuição normal. Sendo assim, pode-se comparar as médias (esperanças matemáticas) dos estimadores para amostras com tempos censurados em relação às com tempos não censurados, mediante os procedimentos baseados na distribuição normal.

O teorema central do limite é enunciado e demonstrado, conforme Spiegel (1978), a continuação.

Teorema 2: Teorema central do limite

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas (todas têm a mesma função de probabilidade, no caso discreto e a mesma função de densidade, no caso contínuo) e com média μ e variância σ^2 finitas. Então, se:

$$S_n = X_1 + X_2 + \dots + X_n, (n = 1, 2, \dots);$$

$$\lim_{n \rightarrow \infty} P(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2}} du$$

isto é, a variável aleatória $\frac{(S_n - n\mu)}{\sigma\sqrt{n}}$, que é variável padronizada correspondente a S_n , é assintoticamente normal. Este teorema é válido também sob condições mais gerais, por exemplo, quando $X_1, X_2, \dots$, são variáveis aleatórias independentes com a mesma média e a mesma variância mas não necessariamente com a mesma distribuição.

Demonstração:

Para $n = 1, 2, \dots$, tem-se $S_n = X_1 + X_2 + \dots + X_n$. Sendo que $X_1, X_2, \dots, X_n$ têm, cada uma, média μ e variância σ^2 . Assim

$$E(S_n) = E(X_1) + E(X_2) + \dots + E(X_n) = n\mu$$

e como os X_k são independentes, então,

$$V(S_n) = V(X_1) + V(X_2) + \dots + V(X_n) = n\sigma^2$$

Segue-se que a variável padronizada correspondente a S_n é:

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

A função geratriz de momentos para S_n^* é

$$\begin{aligned} E(e^{tS_n^*}) &= E[e^{t(S_n - n\mu)/\sigma\sqrt{n}}] = E[e^{t(X_1 - \mu)/\sigma\sqrt{n}}] \dots E[e^{t(X_n - \mu)/\sigma\sqrt{n}}] \\ &= \{E[e^{t(X_1 - \mu)/\sigma\sqrt{n}}]\}^n \end{aligned}$$

em que, nos dois últimos estágios, levamos em conta os fatos que os X_k são independentes e identicamente distribuídas (i.i.d.). Mas, pelo desenvolvimento de Taylor, tem-se:

$$\begin{aligned} E[e^{t(X_1 - \mu)/\sigma\sqrt{n}}] &= E\left[1 + \frac{t(X_1 - \mu)}{\sigma\sqrt{n}} + \frac{t^2(X_1 - \mu)^2}{2\sigma^2 n} + \dots\right] = \\ &= E(1) + \frac{t}{\sigma\sqrt{n}} E(X_1 - \mu) + \frac{t^2}{2\sigma^2 n} E[(X_1 - \mu)^2] + \dots \\ &= 1 + \frac{t}{\sigma\sqrt{n}} (0) + \frac{t^2}{2\sigma^2 n} (\sigma^2) + \dots = 1 + \frac{t^2}{2n} + \dots \end{aligned}$$

de modo que:

$$E(e^{tS_n^*}) = \left(1 + \frac{t^2}{2n} + \dots\right)^n$$

em que o limite dessa expressão, quando $n \rightarrow \infty$ é $e^{\frac{t^2}{2}}$, que é a função geratriz de momentos da distribuição normal padronizada.

O resultado que é desejado decorre do teorema da unicidade, que é enunciado a seguir, porém não é demonstrado.

Teorema 3: Teorema da unicidade

Sejam X e Y variáveis aleatórias com funções geratrizes de momentos $M_X(t)$ e $M_Y(t)$, respectivamente. Então, X e Y terão a mesma distribuição de probabilidade se, e somente se, $M_X(t) = M_Y(t)$.

O Teorema 3 completa a demonstração do teorema central do limite.

A mesma metodologia utilizada para comparar amostras sem censura e com presença de censura foi aplicada para comparar longevidade de machos e fêmeas.

3.6 Teste log-rank

Para comparar dois grupos de indivíduos quanto à sobrevivência, pode-se aplicar o teste não paramétrico log-rank, atualmente o mais utilizado para comparar duas funções de sobrevivência.

A estatística deste teste é obtida de forma similar ao teste de Mantel-Haenszel (1959), para combinar tabelas de contingência.

Para testar as hipóteses:

$$H_0: S_1(t) = S_2(t)$$

$$H_1: S_1(t) \neq S_2(t)$$

considere $t_1 < t_2 < \dots < t_k$, os tempos de falhas distintos da amostra formada pela combinação das duas amostras correspondentes aos dois grupos de indivíduos.

Suponha que no tempo t_j acontecem d_j falhas e n_j indivíduos estão sob risco em um tempo imediatamente anterior a t_j na amostra combinada e, respectivamente, d_{ij} e n_{ij} na amostra i , $i = 1, 2$ e $j = 1, \dots, k$. Os dados podem ser dispostos em uma tabela de contingência, conforme a seguir:

TABELA 2: Tabela de contingência gerada no tempo t_j

	Grupo 1	Grupo 2	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

Se $S_1(t) = S_2(t)$, então $S_1(t_j) = S_2(t_j)$, $\forall j$, e pode-se afirmar que:

$$d_{1j} = d_{2j}, \forall j$$

Também, se $d_{1j} = d_{2j}, \forall j$, então $S_1(t_j) = S_2(t_j), \forall j$

Assim, se $d_{1j} - d_{2j} = 0, \forall j$, então $S_1(t_j) = S_2(t_j), \forall j$

Portanto, a diferença $d_{1j} - d_{2j}$ pode ser usada para testar H_0

$$\begin{cases} \text{Se } d_{1j} - d_{2j} \text{ for pequena, aceitar } H_0 \\ \text{Se } d_{1j} - d_{2j} \text{ for grande, rejeitar } H_0 \end{cases}$$

Fazendo:

$$e_{1j} = \frac{n_{1j}}{n_j} \cdot d_j = \frac{n_{1j}}{n_j} \cdot (d_{1j} + d_{2j})$$

Se d_{2j} aumenta, então e_{1j} aumenta e se d_{2j} diminui, então e_{1j} diminui.

Isto é, existe uma relação diretamente proporcional entre d_{2j} e e_{1j} . Assim:

$(d_{1j} - d_{2j})$ pequeno, então $(d_{1j} - e_{1j})$ pequeno

$(d_{1j} - d_{2j})$ grande, então $(d_{1j} - e_{1j})$ grande

Então, o critério de decisão seria agora:

$$\begin{cases} \text{Se } d_{1j} - e_{1j} \text{ for pequena, } \forall j, \text{ aceitar } H_0 \\ \text{Se } d_{1j} - e_{1j} \text{ for grande, } \forall j, \text{ rejeitar } H_0 \end{cases}$$

Portanto, temos que:

aceitar H_0 , se $\sum_{j=1}^k (d_{1j} - e_{1j})$ for pequeno.

rejeitar H_0 , se $\sum_{j=1}^k (d_{1j} - e_{1j})$ for grande.

Pode-se observar que a distribuição de d_{1j} é hipergeométrica, já que temos um grupo de indivíduos no tempo t_j e é feita uma amostragem sem reposição. Sendo d_{1j} a variável aleatória que indica o número de falhas nesse grupo no tempo t_j , então:

$$P(d_{1j}) = \frac{C_{n_{1j}, d_{1j}} \times C_{n_{2j}, d_{2j}}}{C_{n_j, d_j}} ; d_{1j} = 0, 1, 2, \dots, n_{1j}$$

A média de d_{1j} é $e_{1j} = n_{1j}d_j n_j^{-1}$. Isto quer dizer que, se não houver diferença entre as populações no tempo t_j , 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_{1j} é, então:

$$(V_j)_1 = n_{1j}(n_j - n_{1j})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}$$

Portanto, a estatística $d_{1j} - e_{1j}$ tem média 0 e variância $(V_j)_1$.

Se as k tabelas de contingência forem independentes, um teste aproximado para a igualdade das duas funções de sobrevivência pode ser baseado na estatística

$$T = \frac{\left[\sum_{j=1}^k (d_{1j} - e_{1j}) \right]^2}{\sum_{j=1}^k (V_j)_1}$$

que segue uma distribuição assintoticamente χ^2 com 1 grau de liberdade, para grandes amostras, como demonstrado a seguir:

Sabe-se que para n grande, $\frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}}$ segue assintoticamente uma

distribuição $N(0,1)$.

$$\Rightarrow \left\{ \frac{\frac{\sum_{i=1}^n X_i}{n} - E\left(\frac{\sum_{i=1}^n X_i}{n}\right)}{\sqrt{\text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right)}} \right\}^2 \text{ tem assintoticamente uma } \chi^2_{(1)}.$$

$$\Rightarrow \frac{1/n^2 \left\{ \sum_{i=1}^n [X_i - E(X_i)] \right\}^2}{1/n^2 \left[\sum_{i=1}^n \text{Var}(X_i) \right]} \text{ tem assintoticamente uma } \chi^2_{(1)}.$$

$$\therefore \frac{\left\{ \sum_{i=1}^n [X_i - E(X_i)] \right\}^2}{\left[\sum_{i=1}^n \text{Var}(X_i) \right]}, \text{ tem assintoticamente uma } \chi^2_{(1)}. \quad (I)$$

Então relacionando a expressão (I) com a estatística T , tem-se:

$$r = n$$

$$d_{ij} = X_i$$

$$E(d_{ij}) = e_{ij} = E(X_i)$$

$$\text{Var}(d_{ij}) = \text{Var}(X_i)$$

Assim, tem-se que T tem, assintoticamente, uma distribuição $\chi^2_{(1)}$.

Uma generalização deste teste para comparar mais que duas curvas de sobrevivência pode ser encontrada em Colosimo (2001).

O teste log-rank foi utilizado para comparar curvas de sobrevivência de machos e fêmeas com dados censurados. Primeiro, o teste aplicou-se aos dados reais, obtidos em Silva (1991) e, posteriormente, às 5.000 amostras obtidas por simulação.

Para avaliar a eficiência do teste log-rank foram tomados como referência os resultados obtidos via TCL para testes de hipóteses, segundo o explicado na seção 3.5. Isto é, comparou-se o complemento do nível de significância do teste de hipóteses via TCL com a proporção de testes log-rank que davam significativos entre as 5.000 amostras simuladas.

Como neste trabalho foi possível determinar 5.000 estimativas dos estimadores dos parâmetros para cada grupo, tanto em amostras com presença de censura, como em amostras com tempos de vida completos, tivemos como último objetivo propor e determinar um fator de correção para estimativas do estimador do parâmetro quando os dados são censurados.

3.7 Fator de correção para estimativas censuradas

Após determinadas as estimativas dos parâmetros da distribuição escolhida para as 5.000 amostras simuladas, com censura e sem censura, foram calculadas de forma acumulativa as médias das estimativas, adicionando-as uma a uma, gerando-se um total de 5.000 médias. Para cada média gerada, calculou-se a diferença entre os casos com censura e os casos sem censura.

Com isso, construiu-se um gráfico de dispersão, no qual o número de amostras é a abscissa e a diferença das médias é a ordenada. Assim, encontrou-se um valor d_0 , que representa o valor numérico no qual as diferenças convergem. Matematicamente, esse fator de correção pode ser expresso como:

$$\bar{\lambda}_{sc} = d_0 + \bar{\lambda}_c$$

em que,

$\bar{\hat{\lambda}}_{sc}$ é a média acumulada das estimativas das amostras simuladas sem censura,

$\bar{\hat{\lambda}}_c$ é a média acumulada das estimativas das amostras simuladas com censura.

Então,

$$d_0 = \bar{\hat{\lambda}}_{sc} - \bar{\hat{\lambda}}_c$$

assim, tem-se que d_0 é uma porcentagem de $\bar{\hat{\lambda}}_c$. Logo, o fator de correção para as estimativas censuradas, (F_c), pode ser determinado por:

$$F_c = \frac{d_0}{\bar{\hat{\lambda}}_c} \cdot 100$$

Isto é,

$$F_c = \frac{\bar{\hat{\lambda}}_{sc} - \bar{\hat{\lambda}}_c}{\bar{\hat{\lambda}}_c} \cdot 100.$$

4 RESULTADOS E DISCUSSÃO

4.1 Adequação do modelo

As Figuras 1, 2, 3, 4, 5, e 6 mostram que pelo método dos gráficos de linearização, apresentados em 3.2.1, é possível perceber que os tempos de vidas de machos e fêmeas, seguem o modelo probabilístico Weibull.

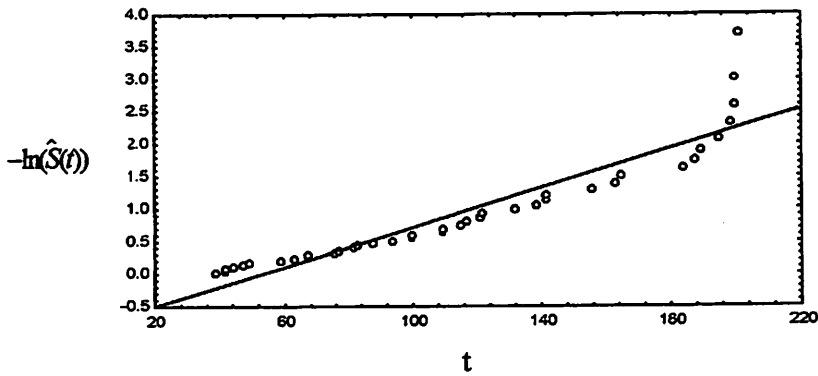


FIGURA 1 Ajuste exponencial para machos

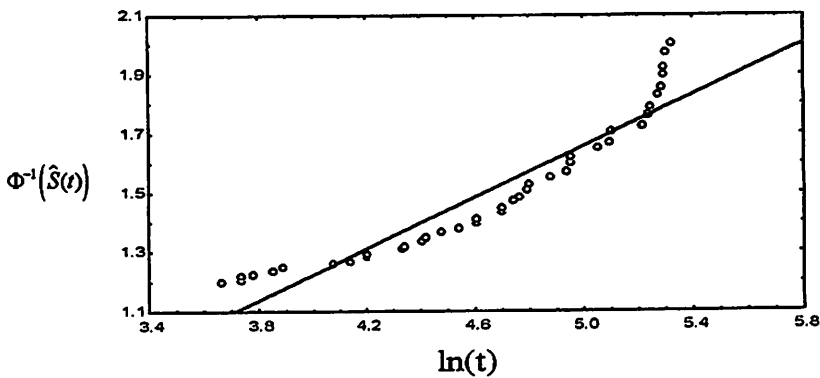


FIGURA 2 Ajuste log-normal para machos

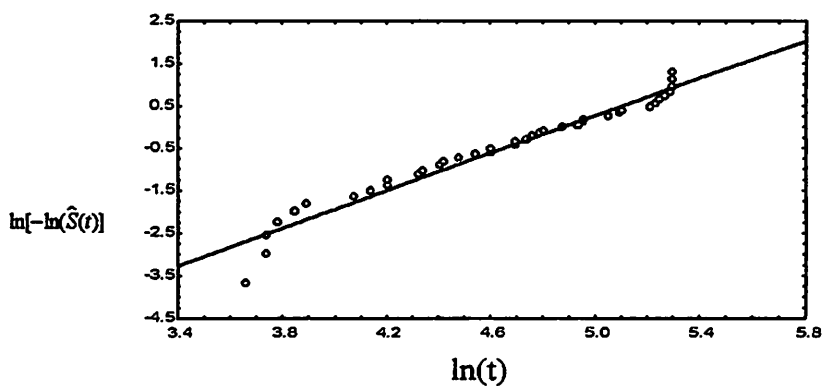


FIGURA 3 Ajuste Weibull para machos

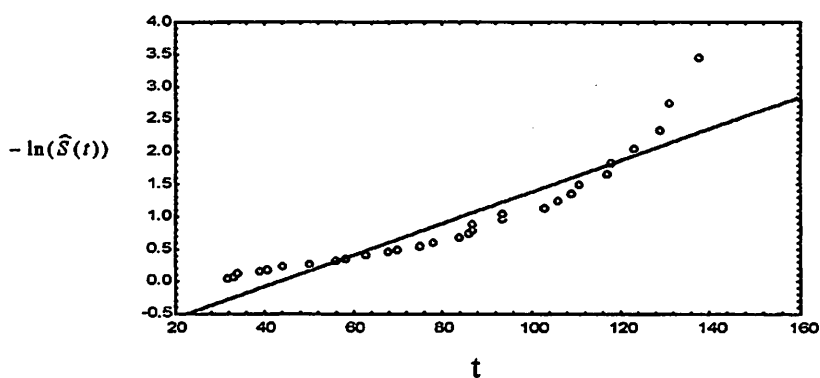


FIGURA 4 Ajuste exponencial para fêmeas

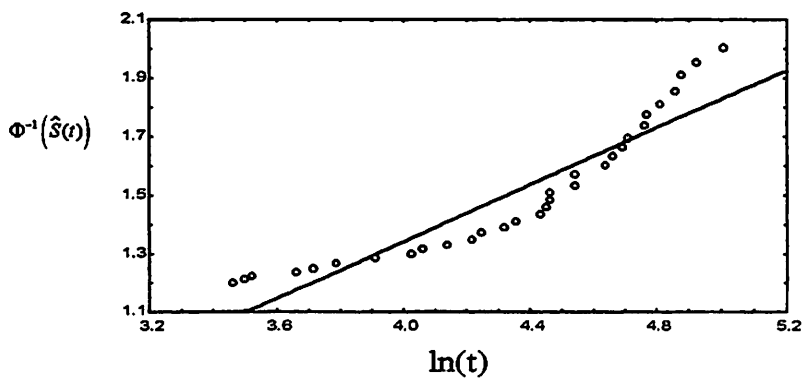


FIGURA 5 Ajuste log-normal para fêmeas

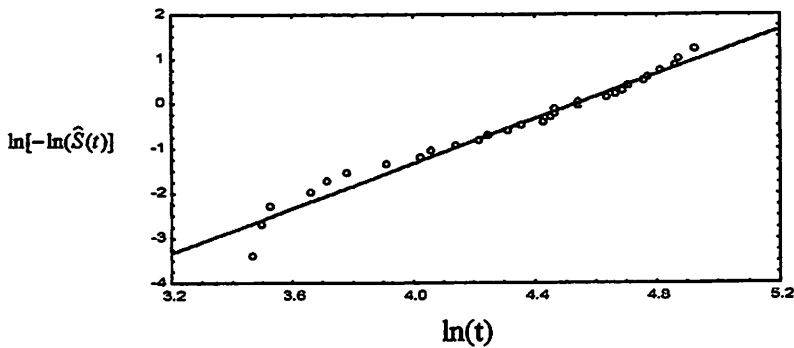


FIGURA 6 Ajuste Weibull para fêmeas

Sendo a distribuição Weibull a que melhor ajusta-se aos dados de longevidade da *Ceraeochrysa cubana* determinados por Silva (1991), com os dados reais estimou-se os parâmetros (α e β) dessa distribuição. Para o conjunto de machos, tem-se $\hat{\alpha}_M = 133,487$ e $\hat{\beta}_M = 2,389$, enquanto que, para fêmeas, $\hat{\alpha}_F = 94,786$ e $\hat{\beta}_F = 2,788$. Assumindo os estimadores de α e β como parâmetros da distribuição Weibull, geraram-se, a partir dessa distribuição, 5.000 amostras para machos e 5.000 para fêmeas, aplicando, posteriormente, em cada amostra, os mecanismos de censura aleatória e tipo I.

4.2 Intervalo de confiança e teste de comparação de médias de estimadores

Abaixo, as tabelas com os respectivos intervalos de confiança, com 95% de confiança, para as médias das diferenças entre as estimações com dados sem censura e com introdução de censura.

21 1.3

TABELA 3 LC. para a média das diferenças entre as estimações com dados sem censura e com 10% de censura

Médias	Censura aleatória	Censura tipo I
$\mu_{\hat{\alpha}_M - \hat{\alpha}_{Mc}}$	(-2,10 ; -1,95)	(-0,704 ; -0,645)
$\mu_{\hat{\beta}_M - \hat{\beta}_{Mc}}$	(-0,0475 ; -0,0416)	(0,0573 ; 0,062)
$\mu_{\hat{\alpha}_F - \hat{\alpha}_{Fc}}$	(-1,085 ; -0,986)	(-1,212 ; -0,667)
$\mu_{\hat{\beta}_F - \hat{\beta}_{Fc}}$	(-0,050 ; -0,042)	(0,0263 ; 0,065)

TABELA 4 LC. para a média das diferenças entre as estimações com dados sem censura e com 20% de censura

Médias	Censura aleatória	Censura tipo I
$\mu_{\hat{\alpha}_M - \hat{\alpha}_{Mc}}$	(-4,51 ; -4,30)	(-0,72 ; 0,61)
$\mu_{\hat{\beta}_M - \hat{\beta}_{Mc}}$	(-0,093 ; -0,084)	(0,0385 ; 0,046)
$\mu_{\hat{\alpha}_F - \hat{\alpha}_{Fc}}$	(-2,298 ; -2,151)	(-1,176 ; -0,637)
$\mu_{\hat{\beta}_F - \hat{\beta}_{Fc}}$	(-0,11 ; -0,097)	(0,015 ; 0,054)

TABELA 5 LC. para a média das diferenças entre as estimações com dados sem censura e com 30% de censura

Médias	Censura aleatória	Censura tipo I
$\mu_{\hat{\alpha}_M - \hat{\alpha}_{Mc}}$	(-7,20 ; -6,92)	(-0,935 ; -0,732)
$\mu_{\hat{\beta}_M - \hat{\beta}_{Mc}}$	(-0,15 ; -0,14)	(0,0245 ; 0,0361)
$\mu_{\hat{\alpha}_F - \hat{\alpha}_{Fc}}$	(-3,75 ; -3,56)	(-0,85 ; -0,654)
$\mu_{\hat{\beta}_F - \hat{\beta}_{Fc}}$	(-0,174 ; -0,158)	(0,026 ; 0,0445)

TABELA 6 I.C. para a média das diferenças entre as estimações com dados sem censura e com 40% de censura

Médias	Censura aleatória	Censura tipo I
$\mu_{\hat{\alpha}_M - \hat{\alpha}_{Mc}}$	(-10,564 ; -10,218)	(-1,212 ; -0,846)
$\mu_{\hat{\beta}_M - \hat{\beta}_{Mc}}$	(-0,223 ; -0,208)	(0,00077 ; 0,017)
$\mu_{\hat{\alpha}_F - \hat{\alpha}_{Fc}}$	(-5,615 ; -5,382)	(-1,34 ; -0,96)
$\mu_{\hat{\beta}_F - \hat{\beta}_{Fc}}$	(-0,264 ; -0,244)	(-0,019 ; 0,0082)

TABELA 7 I.C. para a média das diferenças entre as estimações com dados sem censura e com 50% de censura

Médias	Censura aleatória	Censura tipo I
$\mu_{\hat{\alpha}_M - \hat{\alpha}_{Mc}}$	(-14,91 ; -14,485)	(-1,964 ; -1,332)
$\mu_{\hat{\beta}_M - \hat{\beta}_{Mc}}$	(-0,299 ; -0,281)	(-0,0293; -0,00694)
$\mu_{\hat{\alpha}_F - \hat{\alpha}_{Fc}}$	(-7,696 ; -7,41)	(-2,35 ; -1,575)
$\mu_{\hat{\beta}_F - \hat{\beta}_{Fc}}$	(-0,371 ; -0,345)	(-0,13 ; -0,0885)

Observa-se, nos resultados apresentados nas Tabelas de 3 a 7, que a esperança matemática do estimador para dados sem censura é diferente da esperança matemática do estimador para dados censurados, isto com 95% de confiança. Isto significa que, tanto para machos quanto para fêmeas, a função de sobrevivência encontrada com a introdução de censura é diferente daquela com os dados não censurados. Esta afirmação é feita considerando que a função de sobrevivência representa uma população, da mesma forma que a função de densidade de probabilidade ou a função de distribuição acumulada. Sendo assim, ao ter dois estimadores (por exemplo, $\hat{\alpha}_M$ e $\hat{\alpha}_{Mc}$, para os casos dos machos) com esperanças matemáticas ($\mu_{\hat{\alpha}_M}$ e $\mu_{\hat{\alpha}_{Mc}}$) diferentes, poder-se-ia afirmar que as populações, às quais correspondem tais estimadores, são diferentes, o que implica em funções de sobrevivência diferentes.

Cabe ressaltar que a comparação de populações de dados censurados e dados não censurados, tanto para machos quanto para fêmeas, foi baseada na comparação, de $\mu_{\hat{\alpha}_M}$ versus $\mu_{\hat{\alpha}_{Mc}}$, $\mu_{\hat{\beta}_M}$ versus $\mu_{\hat{\beta}_{Mc}}$, $\mu_{\hat{\alpha}_F}$ versus $\mu_{\hat{\alpha}_{Fc}}$ e $\mu_{\hat{\beta}_F}$ versus $\mu_{\hat{\beta}_{Fc}}$. Todas as comparações foram feitas por meio através de IC e testes de hipóteses baseados na distribuição normal. Este critério foi adotado devido ao fato de com as 5.000 simulações, a média amostral de, $\hat{\alpha}_M$, $\hat{\alpha}_{Mc}$, $\hat{\beta}_M$, $\hat{\beta}_{Mc}$, $\hat{\alpha}_F$, $\hat{\alpha}_{Fc}$, $\hat{\beta}_F$ e $\hat{\beta}_{Fc}$, tem, cada uma, distribuição normal. A base teórica para esta afirmação foi exposta na seção 3.5 quando discutiram-se conceitos relacionados com o teorema central do limite.

Continuando com a discussão das Tabelas 3 a 7, pode-se observar que, tanto para a censura aleatória quanto para censura tipo I, a diferença entre as médias dos estimadores e, portanto, a diferença entre as funções de sobrevivência, aumentam à medida que a porcentagem de dados censurados aumenta. Pode-se concluir, então, que a introdução de censura em dados entomológicos, similares aos utilizados neste trabalho, influencia significativamente na estimação das curvas de sobrevivência.

A Tabela 8 apresenta os resultados dos testes de diferenças de médias, com $\alpha = 0,05$, para comparar as médias dos estimadores de machos *versus* fêmeas, para dados sem a presença de censura. É possível notar que há diferença altamente significativa entre os dois grupos.

TABELA 8 P-value para o teste de diferenças de médias com dados não censurados

ALEATÓRIA		TIPO I	
$H_0 : \mu_{\hat{\alpha}_M} = \mu_{\hat{\alpha}_F}$	$H_0 : \mu_{\hat{\beta}_M} = \mu_{\hat{\beta}_F}$	$H_0 : \mu_{\hat{\alpha}_M} = \mu_{\hat{\alpha}_F}$	$H_0 : \mu_{\hat{\beta}_M} = \mu_{\hat{\beta}_F}$
0,0001	0,0001	0,0001	0,0001

A Tabela 9 apresenta os resultados dos testes de diferenças de médias, com $\alpha = 0,05$, para comparar as médias dos estimadores de machos *versus* fêmeas para dados com introdução de censura. Pode-se observar que todos os resultados são altamente significativos, assim, são iguais aos resultados da Tabela 8. Isto mostra que a metodologia proposta para comparar curvas de sobrevivência, por meio das médias dos estimadores dos parâmetros da distribuição, é adequada.

TABELA 9 P-value para o teste de diferenças de médias com dados censurados

CENSURA	ALEATÓRIA		TIPO I	
	$H_0 : \mu_{\hat{\alpha}_M} = \mu_{\hat{\alpha}_F}$	$H_0 : \mu_{\hat{\beta}_M} = \mu_{\hat{\beta}_F}$	$H_0 : \mu_{\hat{\alpha}_M} = \mu_{\hat{\alpha}_F}$	$H_0 : \mu_{\hat{\beta}_M} = \mu_{\hat{\beta}_F}$
10% censurados	0,0001	0,0001	0,0001	0,0001
20% censurados	0,0001	0,0001	0,0001	0,0001
30% censurados	0,0001	0,0001	0,0001	0,0001
40% censurados	0,0001	0,0001	0,0001	0,0001
50% censurados	0,0001	0,0001	0,0001	0,0001

Em resumo, com um nível de significância de 5%, pode-se notar que todos os testes foram altamente significativos. Ou seja, as populações são diferentes e, com isso, pode-se afirmar que suas funções de sobrevivência são também diferentes.

4.3 Teste log-rank

Na Tabela 10 são apresentados os resultados do teste não-paramétrico log-rank para comparar as curvas de sobrevivência de machos e fêmeas da *Ceraeochrysa cubana*. Neste caso, foram utilizados os dados originais e os resultados que foram conseguidos pelo procedimento PROC LIFETEST do SAS® são apresentados logo a seguir:

TABELA 10 Estatísticas do log-rank para comparação de funções de sobrevivência: machos *versus* fêmeas, dados originais

	ALEATÓRIA		TIPO I	
Completos	11,555	*	11,555	*
10% censurados	9,7865	*	8,8841	*
20% censurados	9,7606	*	6,4702	*
30% censurados	10,3477	*	3,7653	n.s.
40% censurados	9,1188	*	2,9979	n.s.
50% censurados	6,8356	*	2,3559	n.s.

*: existe diferença significativa

n.s.: não existe diferença significativa

As estatísticas apresentadas na Tabela 10 mostram que, mesmo censurando em até 50%, as observações de machos e fêmeas, no caso da censura aleatória, o teste permanece significativo. Já no caso da censura tipo I, a partir da diminuição de 30% do tempo do experimento, as funções de sobrevivência começam a não apresentar diferenças significativas.

Os resultados da Tabela 10 induzem a uma incerteza em relação à eficiência do teste log-rank para comparar curvas de sobrevivência. Tal fato baseia-se no valor da estatística, que diminui à medida que a porcentagem de censura aumenta. Isto é, segundo o log-rank, duas curvas de sobrevivência diferentes são cada vez mais parecidas quando a porcentagem de observações censuradas aumenta. A eficiência do teste log-rank foi medida aplicando-o em cada uma das 5.000 amostras simuladas.

Aplicando o log-rank para cada experimento simulado foi possível determinar a proporção de testes significativos na comparação das funções de sobrevivência de machos *versus* fêmeas. Os resultados são mostrados na Tabela 11.

**TABELA 11 Proporção de testes log-rank significativos ($\alpha = 0,05$):
machos *versus* fêmeas**

	ALEATÓRIA	TIPO I
10% censurados	0,9418	0,9112
20% censurados	0,9324	0,8544
30% censurados	0,9046	0,7600
40% censurados	0,8746	0,6056
50% censurados	0,8436	0,4082

Pode-se observar que, à medida que foi introduzindo porcentagens de censura, aleatória e tipo I, a proporção de testes significativos, foi diminuindo. Nos casos da censura tipo I, esta diminuição começou a ser mais acentuada a partir de 30%, quando iniciou-se a não significância entre as funções de sobrevivência na análise dos dados reais. Além disso, a partir de 30% de censura, o nível de significância obtido por simulação (que é o complemento da proporção de testes significativos) difere muito do nível de significância nominal ($\alpha = 0,05$). Isto mostra que o teste vai perdendo a eficiência com a presença de porcentagens maiores de censura.

4.4 Fator de correção para estimativas censuradas

Utilizando o método proposto na seção 3.7, foram determinados os fatores de correção para as estimativas censuradas, em que as porcentagens de censura tiveram grande influência sobre a estimação e comparação das curvas de sobrevivência. Isto é, para os casos com 40% e 50% de censura. Os resultados são apresentados nas Tabelas 12 e 13 e ilustrados no gráfico de dispersão na Figura 7 (apenas para o caso de machos com 50% de censura aleatória).

TABELA 12 Valores do fator de correção (%)

ALEATÓRIA				
Estimadores	Macho		Fêmea	
	$\hat{\alpha}_c$	$\hat{\beta}_c$	$\hat{\alpha}_c$	$\hat{\beta}_c$
40% censurados	-7,31	-8,15	-5,59	-7,90
50% censurados	-10,00	-10,80	-7,43	-11,24

TABELA 13 Valores do fator de correção (%)

TIPO I				
Estimadores	Macho		Fêmea	
	$\hat{\alpha}_c$	$\hat{\beta}_c$	$\hat{\alpha}_c$	$\hat{\beta}_c$
40% censurados	-0,60	0,12	-1,05	-0,68
50% censurados	-1,11	-0,80	-1,76	-3,97

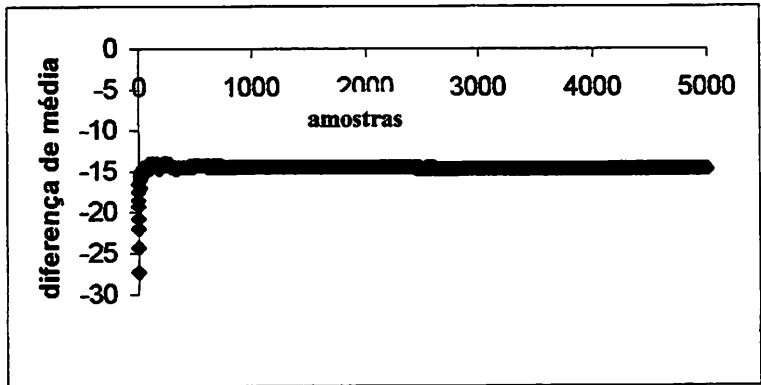


FIGURA 7 Gráfico de dispersão para convergência de d_0 , para o estimador α , para machos, no caso de 50% de censura aleatória.

Com o fator de correção, é possível corrigir empiricamente, em porcentagem, a estimativa para as amostras com a presença de censura. Por exemplo, no caso dos dados para machos com 50% de censura aleatória, o valor

estimado para o parâmetro α da distribuição Weibull é dado por $\hat{\alpha}_{Mc}$. Então, uma aproximação para o valor estimado com os dados sem censura (Tabela 12) seria:

$$\hat{\alpha}_M = \hat{\alpha}_{Mc} - 10\% \hat{\alpha}_{Mc}$$

5 CONCLUSÕES

- 1) A longevidade de machos e fêmeas da *Ceraeochrysa cubana*, determinada por Silva (1991), em ambiente laboratorial segue uma distribuição Weibull.
- 2) O teste log-rank, aplicado à longevidade com informação completa sobre os indivíduos, mostrou que as funções de sobrevivência de machos e fêmeas, são diferentes.
- 3) À medida que se introduz uma porcentagem maior de censura ao conjunto de dados, percebeu-se, via simulação, que aumenta a possibilidade do teste log-rank apresentar resultados incorretos.
- 4) Os intervalos de confiança das médias das diferenças entre as estimações com dados não censurados e as com introdução de censura mostram que a presença de censura faz com que as curvas de sobrevivência resultantes se tornem diferentes das curvas originais, que têm informação completa sobre o tempo de vida dos indivíduos..
- 5) O resultado do teste de diferenças de médias dos estimadores não é influenciado com a introdução de censuras, aleatória e tipo I, nas amostras. Portanto, a aplicação deste teste é mais indicada para esse estudo proposto.
- 6) O fator de correção para as estimativas é indicado, principalmente no caso da introdução de censura aleatória, com 40% e 50% de presença de censura, sendo aqui que orde a influência sobre as estimações é mais crítica.
- 7) É possível aplicar os mecanismos de censura com uma certa restrição, já que a partir da introdução de 30% de censura, tanto na aleatória quanto na tipo I, o principal teste estatístico para a Análise de Sobrevivência passa a não ser confiável.

- 8) Uma sugestão importante é fazer um estudo comparativo utilizando o caso da censura intervalar para verificar se são encontrados os mesmos resultados, já que o estudo realizado por Silva (1991) se assemelha ao caso citado em Colosimo (2001), como exemplo de censura intervalar.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLISON, P. D. **Survival analysis using the SAS system: a practical guide**. Cary, NC: SAS Institute, 1995. 292p.

BOLFARINE, H.; RODRIGUES, J. E.; ACHCAR, J. Análise de Sobrevivência. In: ESCOLA DE MODELOS DE REGRESSÃO, 2, 1991, Rio de Janeiro. **Minicursos...** Rio de Janeiro: Instituto de Matemática da Universidade Federal do Rio de Janeiro, 1991. 111p. Apostila

BRESLOW, N. A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. **Biometrika**, Cambridge, v. 57, n. 3, p. 570-594, Dec. 1970.

CARNEIRO, J. C. Aspectos de confiabilidade de equipamentos elétricos de alta tensão constatados na inspeção e ensaios de recebimento em fábrica: Subsídios para a avaliação da qualidade do produto e suas consequências para o fornecimento de energia elétrica. In: SEMINÁRIO NACIONAL DE PRODUÇÃO E TRANSMISSÃO DE ENERGIA ELÉTRICA, 15, (SNPTEE). GRUPO DE ESUDO DE SUBESTAÇÕES E EQUIPAMENTOS ELÉTRICOS (GSE). 1999, Foz do Iguaçu. Disponível em http://www.itaipu.gov.br/cd_snptee/grupoviii/grupoviii.htm. Acesso em: 20 nov. 2001.

CHALITA, L. V. A. S. **Inferência Paramétrica e Não-Paramétrica em Análise de Sobrevivência**. 1992. 96p. Dissertação (Mestrado em Estatística e Experimentação Agronômica). Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

CHALITA, L. V. A. S. **Modelos para Dados Agrupados e Censurados**. 1997. 135p. Tese (Doutorado em Estatística e Experimentação Agronômica). Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

COLOSIMO, E. A. . Análise de Sobrevivência Aplicada. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGROPECUÁRIA, 9, 2001. Piracicaba. **Minicursos...** Piracicaba: ESALQ/USP, 2001.145p. Apostila.

COX, D. R. Some sample approximate tests for Poisson variates. **Biometrika**, Cambridge, v. 40, n. 3-4, p. 354-360, Dec. 1953.

COX, D. R.; OAKES, D. **Analysis of Survival Data**. London: Chapman and Hall, 1984.195p.

ELANDT-JOHNSON, R. C. E JOHNSON, N. L. **Survival Models and Data Analysis**. New York : John Wiley and Sons, 1980. 457p.

FLEHINGER, B. J., REISER, B., YASHCHIN, E. Survival with competing risks and masked causes of failures. **Biometrika**, Cambridge, v. 85, n. 1, p. 151-164, Mar.1998.

FREITAS, M. A. E COLOSIMO, E. A. **Confiabilidade: Análise de Tempo de Falha e Testes de Vida Acelerados**. Belo Horizonte: Fundação Christiano Ottoni, Escola de Engenharia da UFMG, 1997. 326p.(Série Ferramentas da Qualidade, 12).

GAMEL J. W., WELLER E. A, WESLEY M. N., FERRUR E. J. Parametric cure models of relative and cause-specific survival for grouped survival times. **Computer Methods and Programs in Biomedicine**, Limerick, v. 61, n. 2, p. 99-110, Feb. 2000.

GEHAN, E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. **Biometrika**, Cambridge, v. 52, n. 1 e 2, p. 203-223, June.1965.

GONÇALVES, L. **Biologia e Capacidade Predatória de Podisus nigrolimbatus Spínola, 1832 e Podisus connexivus Bergroth, 1891 (Hemiptera: Pentatomidae: asopinae) em condições de laboratório**. 1990. 97p. Dissertação (Mestrado em Entomologia).Universidade Federal de Lavras, Lavras.

KAPLAN, E. L. E. MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, Washington, v. 53, n. 282, p. 457-481, June. 1958.

LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. New York: John Wiley, 1982. 580 p.

MANTEL, N.; HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. **Journal of the National Cancer Institute**, Washington, v. 22, p. 719-748. 1959.

MICHELETTI, S. M. F. B. Estudos sobre a *Cerconota anonella* (SEEO., 1830) (Lep.: Oecophoridae) em gravioleira (*Annona muricata* L.) no estado de Alagoas. 1999. 87p. Tese (Doutorado em Entomologia). Escola Superior de Agricultura de Luiz de Queiroz, Piracicaba.

MIRAMONTES, O. de; SOUZA, O. The Nonlinear Dynamics of Survival and Social Facilitation in Termites. *Journal of Theoretical Biology*. London, v.181, n. 4, p. 373-380, Aug. 1996

MOOD, A. M.; GRAYBILL, F. A; BOES, D. C. *Introduction to theory of statistics*. 3. Ed. New York: Wiley & Sons, 1974. 842 p.

NASCIMENTO, M. L. Efeito de *Bacillus thuringiensis* var. *kurstaki* sobre *Podisus nigrispinus* (Dallas, 1851) (Heteroptera : Pentatomidae: Asopinae) em Laboratório. 1996. 56p. Dissertação (Mestrado em Entomologia). Universidade Federal de Lavras, Lavras.

OLIVEIRA, L. A. de. *Teste do tipo wilcoxon para a comparação de curvas de sobrevivência*. São Carlos: USP Instituto de Ciências Matemáticas de São Carlos. Instituto de Física e Química de São Carlos. Escola de Engenharia de São Carlos, 1981. 114 p.

RIBEIRO, M. J. *Biologia de Chrysoperla externa* (Hagen, 1861) (Neuroptera, Chrysopidae) alimentada com diferentes dietas. 1988. 131 p. Dissertação (Mestrado em Entomologia). Universidade Federal de Lavras, Lavras.

SANTOS, A. C. dos *Definição do tamanho amostral usando simulação Monte Carlo para os testes de normalidade univariado multivariado baseado s em assimetria e curtose*. 2001, 71 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras, Lavras.

SILVA, R. L. Aspectos bioecológicos e determinação das exigências térmicas de *Ceracochrysa cubana* (Hagen, 1861) (Neuroptera, Chrysopidae) em laboratório. 1991.160 p. Dissertação (Mestrado em Entomologia). Universidade Federal de Lavras, Lavras.

SOARES, J. F.; COLOSIMO, E. Métodos estatísticos na pesquisa clínica. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 40, SIMPÓSIO DE ESTATÍSTICA APLICADA A EXPERIMENTAÇÃO AGRONÔMICA 6, 1995, Ribeirão Preto. *Minicursos...* Ribeirão Preto: RBRAS/USP/UNESP, 1995. 105p. Apostila.

SOUZA, B. de. Aspectos morfológicos e biológicos de nusalala uruguaya (Navás, 1923) (Neuroptera, Hemerobidae) em condições de laboratório. 1988. 195 p. Dissertação (Mestrado em Entomologia). Universidade Federal de Lavras, Lavras.

SPIEGEL, M. R. Probabilidade e Estatística; tradução de: Alfredo Alves Farias. São Paulo: McGraw-Hill do Brasil, 1978. 519p.

TEIXEIRA, J. C., MONTI C. R., DERCHAIN S.. F. M., TEIXEIRA L. C. Avaliação da radioterapia de fatores prognósticos em pacientes com carcinoma epidermóide do colo uterino estágio IIIB. ACTA Oncológica Brasileira, São Paulo, v. 20, n. 1, jan./fev./mar. 2000. Disponível em: <<http://www.cirurgiareparadora.com.br/acta/acta00.html>>. Acesso em: 20 nov. 2001.

VALENÇA, D. M. O modelo de regressão gama generalizada para discriminar entre modelos paramétricos de tempo de vida. 1994. 134 p. (Dissertação – Mestrado em Estatística).Universidade Estadual de Campinas,Campinas.

VIVANCO, M. J. F. Análise de valores extremos no tratamento estatístico da corrosão de equipamentos. 1994. 107 p. Dissertação (Mestrado em Estatística). Universidade Estadual de Campinas,Campinas.

WEIBULL, W. A statistical distribution of wide applicability. Journal of Applied Mechanics, New York, v.18, n.3, p. 293-297, Sep.1951.

ANEXOS

ANEXO

ANEXO.....PÁGINA

SIMULAÇÃO 1 Programa de simulação para gerar dados Weibull via método Monte Carlo, por meio de recursos computacionais do sistema estatístico SAS, introduzindo 50% de censura aleatória.

SIMULAÇÃO 2 Programa de simulação para gerar dados Weibull via método Monte Carlo, por meio de recursos computacionais do sistema estatístico SAS, introduzindo 50% de censura tipo I.

SIMULAÇÃO 1: Programa de simulação para gerar dados Weibull via método de Monte Carlo por meio de recursos computacionais do sistema estatístico SAS, introduzindo 50% de censura aleatória.

*****Cálculo da estimativa de máxima verossimilhança para os parâmetros da Weibull, para machos*****

```
options ps=500 ls=76 nodate number;
data inset1;
  input dias;
  cards;
4.60517
5.10595
4.34381
4.54329
.
.
.
.
;
proc iml worksizes=60;
  use inset1;
  read all into v;
  b0=0.41;
  dr=1;
  do until (dr<0.001);
    n=nrow(v);
    w=j(n,1,1);
    v2=v##2;
    e=exp(v/b0);
    se=w*e;
    d1=(v*e)/se;
    fb0=d1-b0-(w*v)/n;
    d2=(v2*e)/se;
    dfb0=(1/(b0**2))*((d1**2)-d2)-1;
    emvb=b0-fb0/dfb0;
    dr=abs(emvb-b0)/b0;
    b0=emvb;
  end;
  emvb=b0;
  emva=emvb*log((w*exp(v/emvb))/n);
  a=exp(emva);
  b=1/emvb;
  *print emva emvb a b;
```

*****Aplicação do Teorema da Transformação Integral da Probabilidade*****

```
nexp=5000;
n=40;
create sobrevar { dwb, wb, amostra};
```

```

amostra=0;
do exp=1 to nex;
  do i=1 to n;
    u=ranuni(0);
    dwb=((-a**b)*log(1-u))**(1/b);
    amostra= amostra+1;
    append var {dwb amostra};
  end;
  amostra=0;
end;
run;
/*proc print data=sobrev;
var amostra dwb;
quit;*/

```

*****Mecanismo de introdução de censura aleatória: Através do qual há um sorteio de 20 indivíduos a serem censurados e em seguida sorteia-se o tempo de censura para esses indivíduos*****

```

proc iml;
create resobrev var {exp ii dwb wbc cenwb cenwc trat lnwbc lndwb};
use sobrev (keep=dwb);
nexp=5000;
n=40;n1=20;
s1=j(n,1);
do exp=1 to nex;
  read next 40 into x;
  do ii=1 to n;
    s1[ii]=ii;
  end;
  z={0};
  do ii=1 to n1;
    p1=ranuni(0);
    jj=int(p1*(n-ii+1))+1;
    *if jj=0 then jj=1;
    z=z||s1[jj];
    do jjj=1 to n -ii + 1;
      if jj<jjj then s1[jjj-1]=s1[jjj];
      else if jj>jjj then s1[jjj]=s1[jjj];
    end;
  end;
  z=z[2:n1+1];
*print z;
y=x;
do ii=1 to n1;
  jj=z[ii];
  x1=x[jj];
  p2=ranuni(0);
  wbc=p2*x1;
  y[jj]=wbc;
  *x[jj]=y[jj];
end;

```

```

end;
*print y;
do ii=1 to n;
  x1=x[ii];
  x2=y[ii];
  dwb=x1;
  wbc=x2;
  lnwbc=log(x2);
  lndwb=log(x1);
  cenwb=1;
  if dwb > wbc
    then cenwc=0;
  if dwb=wbc then cenwc=1;
  trat=1;
  append var {exp,ii,dwb,wbc,cenwb,cenwc,trat,lnwbc,lndwb};
end;
end;
run;quit;
/*
proc print data=resobrev;
  var exp dwb cenwb wbc cenwc;
run;
quit;*/

```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança para as amostras com censura*****

```

proc lifereg data=resobrev noprint outest=rbrevlr;
model wbc*cenwc(0) = / dist = weibull;
by exp;
run;
/*proc means data=rbrevlr;
var intercep _SCALE_;
run;*/
/*
proc print data=rbrevlr;
var intercep _scale_;
run;quit;*/

```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança com as amostras sem a presença de censura *****

```

proc iml;
  create distem var {exp emva emvb a b };
  use resobrev (keep=lndwb);
  nexp=5000;
  n=40;
  do exp=1 to nexp;
    read next 40 into v;
    b0=0.41;
    dr=1;

```

```

do until (dr<0.001);
  n=nrow(v);
  w=j(n,1,1);
  v2=v##2;
  e=exp(v/b0);
  se=w*e;
  d1=(v*e)/se;
  fb0=d1-b0-(w*v)/n;
  d2=(v2*e)/se;
  dfb0=(1/(b0**2))*((d1**2)-d2)-1;
  emvb=b0-fb0/dfb0;
  dr=abs(emvb-b0)/b0;
  b0=emvb;
end;
emvb=b0;
emva=emvb*log((w*exp(v/emvb))/n);
a=exp(emva);
b=1/emvb;
  append var{exp,emva,emvb,a,b};
*print exp emva emvb a b;
end;
run;quit;
/*
proc print data=distem;
run;quit;*/

```

*****Procedimento para agrupar as estimativas das amostras com e sem censura em um único arquivo de dados ******

```

proc sort data=rbrevlr; by exp;
  proc sort data=distem; by exp;
  data surv;
  merge rbrevlr distem;
  by exp;
  ac=exp(intercep);
  bc=1/_scale_;
  da=a-ac;
  db=b-bc;
run;
proc print data=surv;
var exp a b ac bc;
run;quit;
/*proc univariate data=distem normal;
var a b;*/
* histogram emva emvb a b/normal;
run;quit;

```

*****Cálculo da estimativa de máxima verossimilhança para os parâmetros da Weibull, para fêmeas*****

```
data inseto2;
input dias ;
cards;
4.70953
4.06044
4.46591
.
.
.
.
;
proc iml worksizes=60;
use inseto2;
read all into v;
g0=0.41;
dr=1;
do until (dr<0.001);
n=nrow(v);
w=j(n,1,1);
v2=v##2;
e=exp(v/g0);
se=w*e;
d1=(v*e)/se;
fg0=d1-g0-(w*v)/n;
d2=(v2*e)/se;
dfg0=(1/(g0**2))*((d1**2)-d2)-1;
emvg=g0-fg0/dfg0;
dr=abs(emvg-g0)/g0;
g0=emvg;
end;
emvg=g0;
emvl=emvg*log((w*exp(v/emvg))/n);
l=exp(emvl);
g=1/emvg;
*print emvl emvg l g;
```

*****Aplicação do Teorema da Transformação Integral da Probabilidade*****

```
nexp=5000;
n=31;
create sobrevf var { dwbf, wbf, amostra};
amostra=0;
do exp=1 to nexp;
do i=1 to n;
u=ranuni(0);
dwbf=(-l**g)*log(1-u)**(1/g);
amostra= amostra+1;
append var {dwbf amostra};
end;
```

```

amostra=0;
end;
run;
/*proc print data=sobrevf;
var amostra dwbf;
quit;*/

```

*****Mecanismo de introdução de censura aleatória: Através do qual há um sorteio de 15 indivíduos a serem censurados e em seguida sorteia-se o tempo de censura para esses indivíduos*****

```

proc iml;
create resobref var {exp ii dwbf wbc cenwfb cenwc trat lnwbfc lndwfb};
use sobrevf (keep=dwbf);
nexp=5000;
n=31;n1=15;
s1=j(n,1);
do exp=1 to nexp;
read next 31 into x;
do ii=1 to n;
s1[ii]=ii;
end;
z={0};
do ii=1 to n1;
p1=ranuni(0);
jj=int(p1*(n-ii+1))+1;
*if jj=0 then jj=1;
z=z||s1[jj];
do jjj=1 to n-ii+1;
if jj<jjj then s1[jjj-1]=s1[jjj];
else if jj>jjj then s1[jjj]=s1[jjj];
end;
end;
z=z[2:n1+1];
*print z;
y=x;
do ii=1 to n1;
jj=z[ii];
x1=x[jj];
p2=ranuni(0);
wbc=p2*x1;
y[jj]=wbc;
*x[jj]=y[jj];
end;
*print y;
do ii=1 to n;
x1=x[ii];
x2=y[ii];
dwbf=x1;
wbc=x2;
lnwbfc=log(x2);

```



```

lndwfb=log(x1);
cenwfb=1;
if dwbf<>wbc
    then cenwc=0;
if dwbf=wbc then cenwc=1;
trat=2;
append var {exp,ii,dwbf,wbc,cenwfb,cenwc,trat,lnwbfc,lndwfb};
end;
end;
run;quit;
/*
proc print data=resobref;
var exp dwbf cenwfb wbc cenwc;
run;
quit;*/

```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança para as amostras com censura*****

```

proc lifereg data=resobref noprint outest=rbrevlrf;
model wbc*cenwc(0) = / dist = weibull;
by exp;
run;

```

```

/*proc means data=rbrevlrf;
var intercep _SCALE_;
run;*/
/*
proc print data=rbrevlrf;
var intercep _scale_;
run;quit;*/

```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança com as amostras sem a presença de censura *****

```

proc iml;
create distemf var {exp emvl emvg l g };
use resobref (keep=lndwfb);
nexp=5000;
n=31;
do exp=1 to nexp;
read next 31 into v;
g0=0.41;
dr=1;
do until (dr<0.001);
n=nrow(v);
w=j(n,1,1);
v2=v##2;
e=exp(v/g0);
se=w`*e;
d1=(v`*e)/se;

```

```

fg0=d1-g0-(w`v)/n;
d2=(v2`e)/se;
dfg0=(1/(g0**2))*((d1**2)-d2)-1;
emvg=g0-fg0/dfg0;
dr=abs(emvg-g0)/g0;
g0=emvg;
end;
emvg=g0;
emvl=emvg*log((w`exp(v/emvg))/n);
l=exp(emvl);
g=1/emvg;
append var{exp,emvl,emvg,l,g};
*print exp emvl emvg l g;
end;
run;quit;
/*proc print data=distemf;
run;quit;*/

```

*****Procedimento para agrupar as estimativas das amostras com e sem censura em um único arquivo de dados ******

```

proc sort data=rbrevlrf; by exp;
proc sort data=distemf; by exp;
data survf;
merge rbrevlrf distemf;
by exp;
lc=exp(intercep);
gc=1/_scale_;
dl=1-lc;
dg=g-gc;
run;
/*
proc print data=survf;
var exp l g lc gc;
run;quit;*/
/*proc univariate data=distemf normal;
var l g;*/
* histogram emvl emvg l g/normal;
run;quit;

```

*****Procedimento para agrupar as arquivo de dados de machos e fêmeas ******

```

proc sort data=surv; by exp;
proc sort data=survf; by exp;
data survcom;
merge surv survf;
by exp;
run;
/*
proc print data=survcom;
var exp a b l g ac bc lc gc;

```

```
run;quit;*/
proc means data=survcom cfm alpha=0.05 ;
  var ac bc lc gc da db dl dg;
quit;
/*proc univariate data=survcom normal;
  var a l b g ac lc bc gc;
run; quit;*/
```

*****Aplicação do teste log-rank para as curvas de sobrevivência entre machos e fêmeas para cada experimento*****

```
data rank;
set resobrev resobref; by exp;
run;
OPTIONS PS=7500 LS=75;
proc lifetest data=rank notable;
  time wbc*cenwc(0);
  strata trat;
by exp;
run;
quit;
```

SIMULAÇÃO 2: Programa de simulação para gerar dados Weibull via método de Monte Carlo por meio de recursos computacionais do sistema estatístico SAS, introduzindo 50% de censura tipo I.

*****Cálculo da estimativa de máxima verossimilhança para os parâmetros da Weibull, para machos*****

```
options ps=500 ls=76 nodate number;
data inset1;
  input dias;
  cards;
4.60517
5.10595
4.34381
.
.
.
.
.
;
proc iml worksizes=60;
  use inset1;
  read all into v;
  b0=0.41;
  dr=1;
  do until (dr<0.001);
    n=nrow(v);
    w=j(n,1,1);
    v2=v##2;
    e=exp(v/b0);
    se=w*e;
    d1=(v*e)/se;
    fb0=d1-b0-(w*v)/n;
    d2=(v2*e)/se;
    dfb0=(1/(b0**2))*((d1**2)-d2)-1;
    emvb=b0-fb0/dfb0;
    dr=abs(emvb-b0)/b0;
    b0=emvb;
  end;
  emvb=b0;
  emva=emvb*log((w*exp(v/emvb))/n);
  a=exp(emva);
  b=1/emvb;
  *print emva emvb a b;
```

***** Aplicação do Teorema da Transformação Integral da Probabilidade*****

```
nexp=5000;
n=40;
create sobrevar { dwb, amostra, exp};
  amostra=0;
  do exp=1 to nexp;
```

```

do i=1 to n;
  u=ranuni(0);
  dwb=((-a**b)*log(1-u))**(1/b);
  amostra= amostra+1;
  append var {dwb amostra exp};
end;
amostra=0;
end;
run;
proc sort data=sobrev out=ordsobr;
by exp dwb ;
quit;
run;
/* proc print data=ordsobr;
run; quit;*/
proc iml;
create mat2 var {wbc,exp,cenwc,cenwb,trat, lndwb, lnwbc};
use ordsobr (keep=exp dwb);

```

*****Mecanismo de introdução de censura tipo I: Em que verifica-se o maior tempo de vida para o experimento e em seguida o experimento é diminuído em 50% do seu tempo, sendo que os indivíduos com tempo de vida superior a este novo tempo final passam a ser censurados, com um tempo de censura igual a esse tempo final do experimento*****

```

nexp=5000;
do exp=1 to nexp;
read next 40 into valor;
c1d=valor[1:40,1:1];
*print c1d;
aux=c1d[40,1]-0.5*c1d[40,1];
*print aux;
c2d=j(40,1,0);
do i=1 to 40;
  cenwb=1;
  trat=1;
c2d[i,1]=c1d[i,1];
k=c1d[i,1];
if aux<k then
  c2d[i,1]=aux;
  wbc=c2d[i,1];
if aux<k then
  cenwc=0;
else cenwc=1;
lndwb=log(k);
lnwbc=log(wbc);
append var {exp wbc cenwb cenwc trat lndwb lnwbc};
end;
end;
run; quit;
/* proc print data=mat2;
run; quit;*/

```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança para as amostras com censura*****

```
proc lifereg data=mat2 noprint outest=rbrevlr;
model wbc*cenwc(0) = / dist = weibull;
by exp;
run;
/*proc means data=rbrevlr;
var intercep _SCALE_;
run;*/
/*
proc print data=rbrevlr;
var intercep _scale_;
run;quit;*/
```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança com as amostras sem a presença de censura *****

```
proc iml;
create distem var {exp emva emvb a b };
use mat2 (keep=Indwb);
nexp=5000;
n=40;
do exp=1 to nexp;
read next 40 into v;
b0=0.41;
dr=1;
do until (dr<0.001);
n=nrow(v);
w=j(n,1,1);
v2=v##2;
e=exp(v/b0);
se=w*e;
d1=(v*e)/se;
fb0=d1-b0-(w*v)/n;
d2=(v2*e)/se;
dfb0=(1/(b0**2))*((d1**2)-d2)-1;
emvb=b0-fb0/dfb0;
dr=abs(emvb-b0)/b0;
b0=emvb;
end;
emvb=b0;
emva=emvb*log((w*exp(v/emvb))/n);
a=exp(emva);
b=1/emvb;
append var{exp,emva,emvb,a,b};
*print exp emva emvb a b;
end;
run;quit;
/*
proc print data=distem;
```

```
run;quit;*/
```

*****Procedimento para agrupar as estimativas das amostras com e sem censura em um único arquivo de dados ******

```
proc sort data=rbrevlr; by exp;
  proc sort data=distem; by exp;
  data surv;
  merge rbrevlr distem;
  by exp;
  ac=exp(intercep);
  bc=1/_scale_;
  da=a-ac;
  db=b-bc;
run;
/*
proc print data=surv;
var exp a b ac bc;
run;quit;*/
/*proc univariate data=distem normal;
var a b;*/
* histogram emva emvb a b/normal;
run;quit;
/*proc univariate data=surv normal;
var dia dib;
run; quit;*/
/*proc means data=surv clm alpha=0.05;
var da db;
run;quit;*/
```

*****Cálculo da estimativa de máxima verossimilhança para os parâmetros da Weibull, para fêmeas*****

```
data inseto2;
input dias ;
cards;
4.70953
4.06044
4.46591
.
.
.
.
;
proc iml worksizes=60;
use inseto2;
read all into v;
g0=0.41;
dr=1;
do until (dr<0.001);
  n=nrow(v);
  w=j(n,1,1);
```

```

v2=v##2;
e=exp(v/g0);
se=w`*e;
d1=(v`*e)/se;
fg0=d1-g0-(w`*v)/n;
d2=(v2`*e)/se;
dfg0=(1/(g0**2))*((d1**2)-d2)-1;
emvg=g0-fg0/dfg0;
dr=abs(emvg-g0)/g0;
g0=emvg;
end;
emvg=g0;
emvl=emvg*log((w`*exp(v/emvg))/n);
l=exp(emvl);
g=1/emvg;
*print emvl emvg l g;

```

*** Aplicação do Teorema da Transformação Integral da Probabilidade***

```

nexp=5000;
n=31;
create sobrevf var {dwbf, amostra,exp};
amostra=0;
do exp=1 to nexp;
do i=1 to n;
u=ranuni(0);
dwbf=(-l**g)*log(1-u)**(1/g);
amostra= amostra+1;
append var {dwbf amostra exp};
end;
amostra=0;
end;
run;
/*proc print data=sobrevf;
var amostra dwbf;
quit;*/
proc sort data=sobrevf out=ordsobf;
by exp dwbf ;
quit;
run;
/* proc print data=ordsobf;
run; quit;*/
proc iml;
create mat3 var {wbc,exp,cenwc,cenwb,trat,lnwfb,lnwbc};
use ordsobf (keep=exp dwbf);

```


*****Mecanismo de introdução de censura tipo I: Em que verifica-se o maior tempo de vida para o experimento e em seguida o experimento é diminuído em 50% do seu tempo, sendo que os indivíduos com tempo de vida superior a este novo tempo final passam a ser censurados, com um tempo de censura igual a esse tempo final do experimento*****

```
nexp=5000;
do exp=1 to nexp;
read next 31 into valor;
*print valor;
c1d=valor[1:31,1:1];
*print c1d;
aux=c1d[31,1]-0.5*c1d[31,1];
*print aux;
c2d=j(31,1,0);
do i=1 to 31;
cenwb=1;
trat=2;
c2d[i,1]=c1d[i,1];
k=c1d[i,1];
if aux<k then
c2d[i,1]=aux;
wbc=c2d[i,1];
if aux<k then
cenwc=0;
else cenwc=1;
lndwfb=log(k);
lnwbc=log(wbc);
append var {exp wbc cenwb cenwc trat lndwfb lnwbc};
end;
end;
run; quit;
/*proc print data=mat3;
run; quit;*/
```

*****Cálculo das estimativas dos estimadores de máxima verossimilhança para as amostras com censura*****

```
proc lifereg data=mat3 noprint outest=rbrevlrf;
model wbc*cenwc(0) = / dist = weibull;
by exp;
run;
/*proc means data=rbrevlrf;
var intercep _SCALE_;
run;*/
/*
proc print data=rbrevlrf;
var intercep _scale_;
run;quit;*/
```

***** Cálculo das estimativas dos estimadores de máxima verossimilhança com as amostras sem a presença de censura *****

```
proc iml;
  create distemf var {exp emvl emvg l g };
  use mat3 (keep=indwfb);
  nexp=5000;
  n=31;
  do exp=1 to nexp;
    read next 31 into v;
    g0=0.41;
    dr=1;
    do until (dr<0.001);
      n=nrow(v);
      w=j(n,1,1);
      v2=v##2;
      e=exp(v/g0);
      se=w*e;
      d1=(v*e)/se;
      fg0=d1-g0-(w*v)/n;
      d2=(v2*e)/se;
      dfg0=(1/(g0**2))*((d1**2)-d2)-1;
      emvg=g0-fg0/dfg0;
      dr=abs(emvg-g0)/g0;
      g0=emvg;
    end;
    emvg=g0;
    emvl=emvg*log((w*exp(v/emvg))/n);
    l=exp(emvl);
    g=1/emvg;
    append var{exp,emvl,emvg,l,g};
    *print exp emvl emvg l g;
  end;
run;quit;
/*proc print data=distemf;
run;quit;*/
```

*****Procedimento para agrupar as estimativas das amostras com e sem censura em um único arquivo de dados ******

```
proc sort data=rbrevlrf; by exp;
proc sort data=distemf; by exp;
data survf;
merge rbrevlrf distemf;
by exp;
lc=exp(intercep);
gc=1/_scale_;
dl=1-lc;
dg=g-gc;
run;
```

```

/*
proc print data=surv;
var exp l g lc gc;
run;quit;*/
/*proc univariate data=distemf normal;
var l g;*/
* histogram emvl emvg l g/normal;
run;quit;
/*proc univariate data=surv normal;
var dil dig;
run; quit;*/
/*proc means data=surv clm alpha=0.05;
var dl dg;
run;quit;*/

```

*****Procedimento para agrupar as arquivo de dados de machos e fêmeas *****

```

proc sort data=surv; by exp;
proc sort data=surv; by exp;
data survcom;
merge surv surv;
by exp;
run;
/*
proc print data=survcom;
var exp ac bc lc gc;
run;quit;*/
proc means data=survcom clm alpha=0.05 ;
var ac bc lc gc da db dl dg;
run;
quit;
/*proc univariate data=survcom normal;
var a l b g ac lc bc gc;
run; quit;*/
data rank;
set mat2 mat3; by exp;
run;
quit;
/*proc print data=rank;
var wbc trat cenwc;
run;*/

```

*****Aplicação do teste log-rank para as curvas de sobrevivência entre machos e fêmeas para cada experimento*****

```

OPTIONS PS=7500 LS=75;
proc lifetest data=rank NOTABLE;
time wbc*cenwc(0);
strata trat;
by exp;
run;quit;

```