



Análise de Sobrevivência aplicada a dados de Colhedoras de cana-de-açúcar.

Bruna Aparecida Wruck Verssani¹, Edwin Moisés Marcos Ortega²

¹ Aluna de Mestrado do PPG em Estatística e Experimentação Agronômica/ ESALQ (USP), Piracicaba, SP, Brasil

² Professor Doutor do Dep. de Ciências Exatas / ESALQ (USP), Piracicaba, SP, Brasil

RESUMO

O estudo do tempo de vida ou durabilidade de um produto é o principal objetivo daqueles que se dedicam a Análise de Sobrevivência. O problema apresentado refere-se ao tempo de vida útil de uma peça importante para as Colhedoras de cana-de-açúcar. Devido à complexidade do sistema e a presença de vários fatores que influenciam nessa durabilidade viu-se a necessidade da inclusão de outras variáveis, ou seja, a análise adequada é através de Modelos de Regressão.

Palabras chave: Análise de Sobrevivência, Confiabilidade, Modelos de Regressão.

1. INTRODUÇÃO

Análise de sobrevivência (ou confiabilidade) é um conjunto de técnicas e modelos estatísticos frequentemente utilizados nas áreas biológicas além de possuir aplicações em diversas áreas do conhecimento, como engenharia, demografia e economia. O principal interesse é modelar a variável resposta “tempo”, ou seja, o tempo até a ocorrência de um evento de interesse. Uma característica importante desses dados é a presença de censura, que é a observação parcial da resposta (Colosimo e Giolo 2006). Pode-se citar como aplicação dessa técnica, o estudo da durabilidade e do risco de falhas de determinadas peças ou máquinas agrícolas.

O Brasil é o maior produtor de cana-de-açúcar, segundo o IBGE (Instituto Brasileiro de Geografia e Estatística), a cana-de-açúcar ocupa cerca 10% de todas as áreas utilizadas pela agricultura brasileira sendo responsável por pouco mais de 17% das exportações nacionais. Daí a importância da qualidade e durabilidade dos equipamentos envolvidos nesse processo.

O problema em estudo refere-se aos dados de colhedoras de cana-de-açúcar de quatro unidades de uma Usina produtora de açúcar e álcool. Na safra de 2014/2015, foram gastos 8.589.230,87 reais com peças de 112 colhedoras. A escolha da peça que será analisada se deu devido aos altos custos envolvidos em sua manutenção, já que os gastos com a Lâmina Chopper corresponderam a 469.784,19 reais, ou seja, 5.5% dos gastos totais com peças.

A modelagem desses dados tem o objetivo encontrar a curva de sobrevivência associada a cada Unidade da Usina, levando em consideração os tipos de Agente Causador da Falha.

2. METODOLOGIA

O desenvolvimento do trabalho consistiu, primeiramente, do estudo do sistema gerador das falhas, coleta e manipulação do banco de dados sendo analisadas todas as colhedoras e peças. As variáveis contidas no banco de dados são Unidade da Usina, Falha (se a peça falhou ou censurou), Durabilidade, Estado e Agente causador da Falha. Neste contexto a censura ocorre quando a peça não falhou durante o período determinado, ou seja, a safra. Sendo todas censura à direita. As análises foram realizadas utilizando o Software R.

O estimador de Kaplan-Meier (KM) é considerado um método padrão para a estimação da função de sobrevivência empírica quando os dados são censurados. O estimador KM não assume nenhuma suposição sobre a distribuição para o tempo, por isso é considerado um estimador não paramétrico. A Figura 1 mostra as curvas de sobrevivência estimada para cada Unidade da Usina. Nota-se que essas curvas têm comportamento semelhante e se cruzam não obedecendo ao pressuposto de proporcionalidade exigido pelo teste Log-Rank, sendo assim não é possível compará-las.

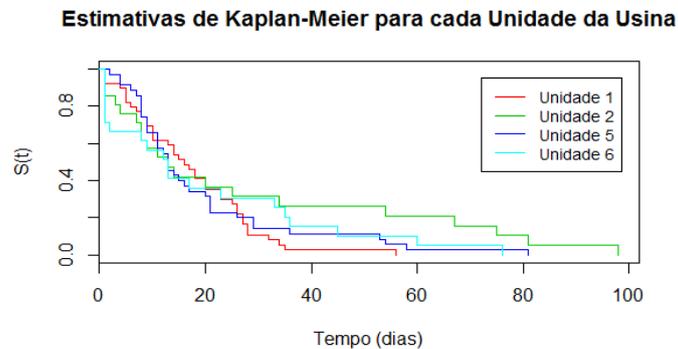


Figura 1 – Curvas estimadas de Sobrevivência para as Unidades da Usina.

A construção do gráfico TTT-Plot dá indicativo do comportamento da função de risco. Assim, testamos as distribuições de probabilidade que sua função de risco tenha comportamento semelhante. A curva próxima à reta nos indica um comportamento constante para a função de risco e a construção do envelope de 90% (Louzada-Neto, F. e Andrade, C. S. 2001) comprova esse indicativo.

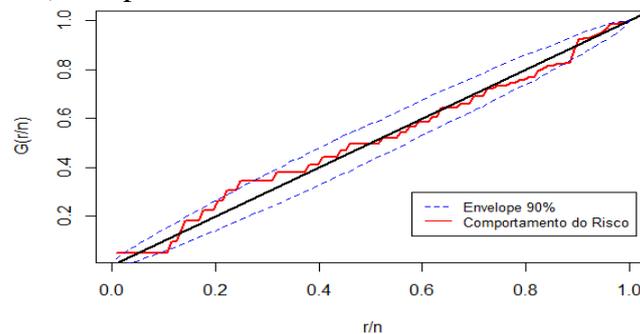


Figura 2 – Gráfico TTT – Plot

Para auxiliar na escolha do modelo desconsiderou as covariáveis e construiu os gráficos das linearizações para os modelos exponencial, weibull e log-normal, respectivamente. Apresentam-se, na Figura 3, como melhores candidatos dentre os citados os modelos Weibull e Exponencial, pois seus pontos estão próximos à reta diagonal.

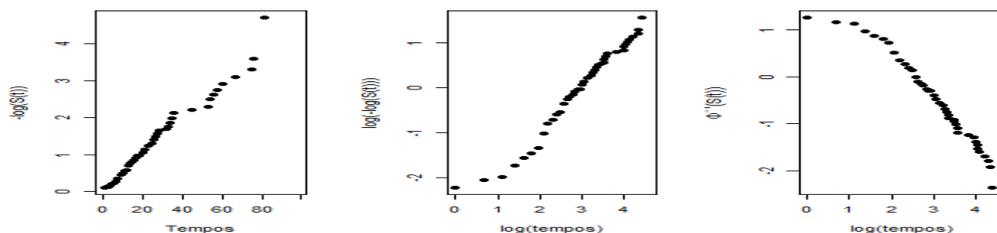


Figura 3 – Gráficos das linearizações.

Os dados apresentam mais de uma covariável, exigindo uma modelagem através de métodos de regressão. A Tabela 1 contém as estimativas dos parâmetros e os valores-p para testar a significância dos parâmetros da Regressão Exponencial e Regressão Weibull (a regressão foi feita no parâmetro de escala).

Tabela 1 – Estimativas para os dados da Lâmina Chopper.

Regressão exponencial		Regressão Weibull	
Estimativas	p-valor	Estimativas	p-valor
$\beta_0 = 3,771$	$9,33e-17$	$\beta_0 = 3,785$	$8,29e-20$
$\beta_1 = 0,010$	0,834	$\beta_1 = 0,014$	0,738
$\beta_2 = -0,385$	0,136	$\beta_2 = -0,410$	0,084
$\beta_3 = -0,536$	0,176	$\beta_3 = -0,513$	0,158
$\gamma = 1$ (fixo)		$\gamma = 1,091$	0,236

Com base no nível de significância de 5% não temos evidência estatística para rejeitar a hipótese de que o efeito dos parâmetros é nulo. Entretanto, devido à importância prática dessas covariáveis para discriminação das Unidades optou-se por mantê-las no modelo.

Testando-se as hipóteses $H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$, obteve-se para o teste de razão de verossimilhança o valor TRV=1,3437 (valor p = 0,246). Este resultado fornece indicações favoráveis ao modelo de regressão exponencial. Entretanto, isso não atesta a adequabilidade do modelo, sendo necessária uma avaliação do modelo ajustado. Para esta avaliação, foram utilizados os resíduos de Cox-Snell, definidos por:

$$\hat{e}_i = [t_i \exp(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3)]$$

Verifica-se na figura 4 que as estimativas das curvas de sobrevivência dos resíduos obtidos por Kaplan-Meier e pelo modelo exponencial padrão estão próximas, e o gráfico dos pares de pontos S(e) – Kaplan-Meier e S(e) – Exponencial padrão estão próximos de uma reta. Assim, podemos considerar um bom ajuste dos dados ao modelo de Regressão Exponencial.

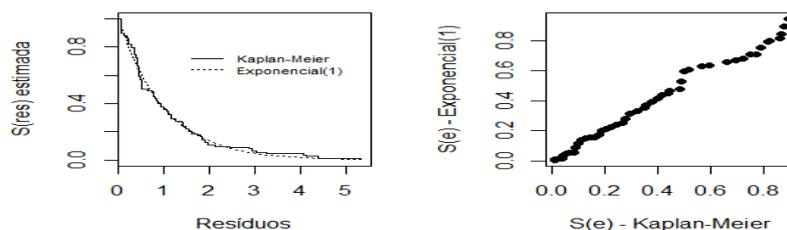


Figura 4 – Análise dos resíduos de Cox Snell do modelo de Regressão Exponencial

A função de sobrevivência obtida pelo modelo de regressão exponencial ajustado aos dados da Lamina Chopper é expressa por:

$$\hat{S}(t | x_1, x_2, x_3) = \exp \left\{ - \left(\frac{t}{\exp(3,771 + 0,010x_1 - 0,385x_2 - 0,536x_3)} \right) \right\}$$

em que x_1 é a Unidade, x_2 é o Estado e x_3 é o agente causador da falha.

3. RESULTADOS E CONCLUSÕES

Com a função de sobrevivência estimada, foi possível encontrar as curvas de sobrevivência (confiabilidade) para cada Unidade da Usina, considerando o tipo de Agente causador da falha e o Estado que cada Unidade pertence.

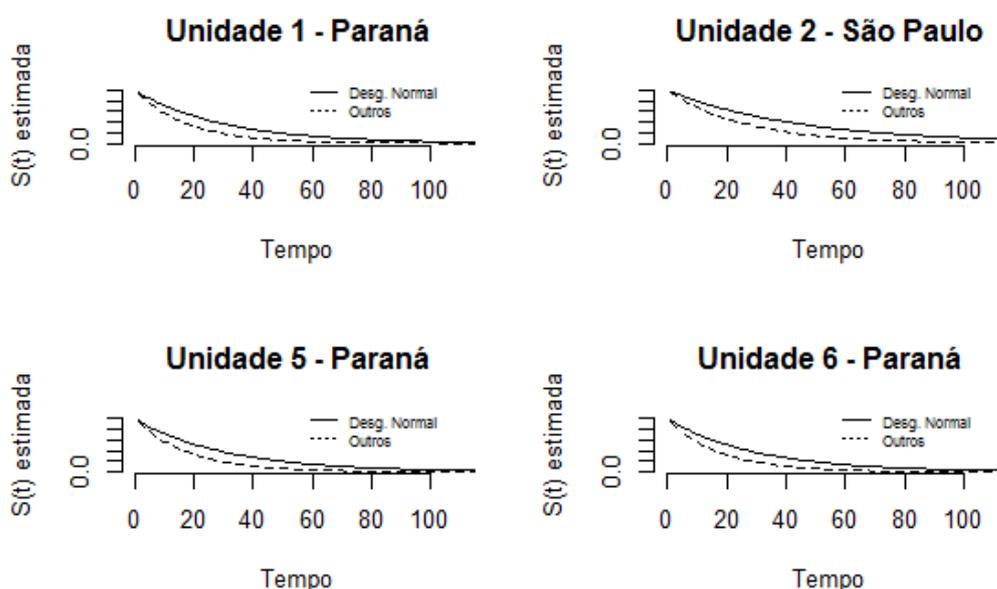


Figura 5 – Curvas de sobrevivência estimadas pelo modelo de regressão exponencial para os dois tipos de Agente Causador da falha.

Têm-se que a sobrevivência é maior para os casos em que houve Desgaste Normal da peça, com comportamento semelhante entre as Unidades. Essa conclusão já era esperada, pois em “Outros” encontram-se aquelas peças que falharam por acidente, manutenção errada ou material deficiente.

Referências

- [1] COLOSIMO, E. A. and Giolo, S. R. Análise de sobrevivência aplicada. Editora Edgard Blücher, 2006.
- [2] IBGE – Instituto Brasileiro de Geografia e Estatística. 2016. Estatística da Produção Agrícola. Disponível em: < ftp://ftp.ibge.gov.br/Producao_Agricola/Fasciculo_Indicadores_IBGE/estProdAgr_201601.pdf > . Acesso em 20 de out 2016.
- [3] LOUZADA-NETO, F., & Andrade, C. S. (2001). Detecting non-Identifiability on the poly-Weibull model. Brazilian Journal of Probability and Statistics, 147-154