Universidade Federal do Paraná

Setor de Ciências Exatas

Departamento de Estatística

Programa de Especialização em *Data Science* e *Big Data*

Ailton Pinto

Análise de sobrevivência de Unidades Hidrogeradoras de Eletricidade

Curitiba

2022

Ailton Pinto

Análise de sobrevivência de Unidades Hidrogeradoras de Eletricidade

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. José Luiz Padilha da Silva

Curitiba

2022

Análise de sobrevivência de Unidades Hidrogeradoras de Eletricidade

Ailton Pinto[1]
Prof. José Luiz Padilha da Silva[2]

Resumo

Em uma análise de sobrevivência de unidades hidrogeradoras de eletricidade realizada em uma empresa de Geração no Estado do Paraná foram obtidos os indicadores de Confiabilidade, tempos de vida médios, análises exploratória de dados, estimativa de parâmetros para uma distribuição de Weibull. O modelo probabilístico obtido pode fornecer melhor entendimento para estimativas de falhas em períodos definidos, predição de intervalos de confiança aplicáveis a otimização e previsibilidade das políticas de manutenção.
Palavras-chave: Análise Sobrevivência, Distribuição Weibull, Confiabilidade, Intervalo de Confiança, Predição, Estimativa Parâmetros, Geração, Hidrogeradores, Eletricidade, Manutenção, Falhas.

Abstract

In a survival analysis of electricity hydrogenerator units carried out in a company that operates Generation in the State of Paraná, reliability indicators, average lifetimes, exploratory data analysis, parameter estimation for a Weibulll distribution were obtained. The probabilistc model obtained can provide a better understanding for estimating failures in defined periods, predicting confidence intervals applicable to optimization and predictability of maintenance policies.
***Keywrods**: Survival Analysis, Weibull Distribution, Reliability, Confidence Interval, Prediction, Estimation, Parameters, Generation, Hydrogenerators, Electricity, Maintenance, Failures.*

1 Introdução

A eletricidade tem papel importante na vida cotidiana, e sua geração é atividade econômica de grande importância para a sociedade brasileira. Dentre as fontes de geração da matriz elétrica brasileira destaca-se a hidráulica que representa 53,4% da demanda total, somando um total de 362TWh de oferta de energia elétrica em 2021[1]. Dada a importância das usinas hidroelétricas, o estudo do comportamento da disponibilidade destes ativos é por consequência bastante relevante. Neste estudo vamos analisar os dados de tempo de falha das unidades geradoras hidroelétricas de uma usina instalada no Rio Iguaçu no Estado do Paraná. Esta usina tem capacidade de geração total de 1.240MW, com 4 unidades geradoras de potência de 310MW cada. O dados disponíveis compreendem o período desde a sua entrada em operação comercial em 1999 até o setembro de 2021. O principal objetivo é identificar um modelo probabilístico para analisar os tempos entre as falhas bem como obter medidas para a confiabilidade dos hidrogeradores.

A NBR 5462[2] define confiabilidade como sendo a capacidade de um item exercer uma função requerida sob condições especificadas durante certo período de tempo, e também a medida de confiabilidade como a função $R(t_1, t_2)$ como a probabilidade de um item poder desempenhar uma função requerida, sob dadas condições, durante um dado intervalo de tempo (t_1, t_2). A função de confiabilidade (ou função de sobrevivência) $R(t)$ indica a probabilidade de uma unidade apresentar sucesso em seu funcionamento (ou seja, ausência de falhas) no intervalo de tempo $(0, t)$ e ainda estar funcionando no tempo t[3]. O estudo da confiabilidade das unidades geradoras pode ser ferramenta para direcionar as políticas de manutenção e a prover parâmetros para indicadores chave da manutenção. A Confiabilidade dos ativos de geração é prioridade para as empresas do setor porque o modelo do sistema elétrico brasileiro é regulado e fiscalizado pela Agência Nacional de Energia Elétrica - ANEEL que define parâmetros de operação dos quais a disponibilidade é um importante componente na receita de geração.

Este trabalho é organizado como se segue: Capítulo 2 Metodologia, Capítulo 3 Análise Exploratória do Dados, Capítulo 4 Análise de Sobrevivência e finalmente no Capítulo 5 Conclusões.

2 Metodologia

Inicialmente os dados precisam ser estudados pela Análise de Exploratória de dados com o uso de medidas resumo: média, mediana, máximo, mínimo, número total e tabelas resumo de frequências das variáveis apresentadas na base de dados. Além disto se faz necessário uma análise gráfica através de boxplot e histograma

[1] Aluno do programa de Especialização em Data Science & Big Data, `ailtonpto@gmail.com`.
[2] Professor do Departamento de Estatística - DEST/UFPR.

para identificação de uma distribuição de probabilidade adequada e análise das premissas de assimetria ou tendências.

2.1 Análise de Sobrevivência

A Análise de Sobrevivência ou Confiabilidade, utiliza um conjunto de técnicas estatísticas para análise de dados de tempo de vida, ou seja, o tempo decorrido entre o tempo inicial e o tempo da ocorrência do evento de interesse.

Pode-se definir a variável aleatória $T \geqslant 0$ como sendo o tempo de sobrevida [4] e [5].

A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t, ou seja, a probabilidade de uma observação sobreviver ao tempo t:

$$S(t) = P(T \geqslant t).$$

A função taxa de falha ou função de risco $h(t)$ no intervalo $[t_1, t_2)$ é expressa por:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T + \Delta t | T \geqslant t)}{\Delta t}.$$

A função de risco acumulada $H(t)$, que é uma função não negativa e monótona crescente, é definida por:

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

ou

$$H(t) = -\log(S(t)).$$

A função acumulada $F(t)$, mede o risco de ocorrência do acontecimento de interesse até o instante t, pode ser definida por:

$$F(t) = P(T \leqslant t), 0 \leqslant t < \infty.$$

2.1.1 Estimador Não-paramétrico de Kaplan-Meier

O estimador não-paramétrico de Kaplan-Meier, proposto Kaplan e Meier(1958) para estimar a função de sobrevivência $S(t)$ é uma adaptação da função de sobrevivência empírica que, na ausência de censura, é definida por:

$\hat{S}(t)$ = (no. de observações que não falharam até o tempo t) / (número total de observações no estudo).

Formalmente considere:

•$t_1 < t_2 \ldots < t_k$, os k tempos distintos e ordenados de falha,

•d_j, o número de falhas em t_j, $j = 1, 2, 3, \ldots k$ e,

•n_j, o número de indivíduos sob risco em t_j, ou seja os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j.

O estimador de Kaplan-Meier é definido por:

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

2.1.2 Distribuição de Weibull

A distribuição de Weibull foi proposta originalmente por W. Weibull(1954) em estudos relacionados ao tempo de falha devido à fadiga de metais. Esta distribuição é bastante popular devido à sua propriedade de apresentar várias formas e ter sua função de falha estritamente crescente, estritamente decrescente ou constante (monótona).

A f.d.p.(função densidade de probabilidade) de uma variável aleatória T com distribuição de *Weibull* é dada por:

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

Dados o parâmetro de forma γ não tem unidade de medida e é positivo. O parâmetro de escala α têm a mesma unidade de medida do tempo e é positivo.

A distribuição de Weibull foi escolhida especificamente neste estudo pois já é amplamente usada na Engenharia de Manutenção, e na empresa já foi utilizada anteriormente para análises de tempos de vida, sendo familiar seu uso e também como o conhecimento de suas propriedades úteis como a interpretação dos modos de falha pelo valor forma γ e escala α relacionado ao tempo de vida.

Para a variável resposta tempo T , e X o vetor de covariáveis:$x_1, \ldots, x_p$, que segue uma distribuição de Weibull, o modelo linear pode ser escrito:

$$Y = log(T) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p = X\beta + \sigma\nu.$$

A função de sobrevivência $S(t|x)$ é dados por:

$$S(t|x) = \exp\left\{ -\left(\frac{t}{\exp\{x\beta\}} \right)^{1/\sigma} \right\}.$$

2.1.3 Função de Verossimilhança

Dado um parâmetro genérico θ, e suponha uma amostra de observações $t_1, \ldots, t_n$, e a variável indicadora δi é 0 = Censura e 1 = Ocorrência do evento, com o parâmetro θ para distribuição Weibull é igual ao par (α, γ), então a função de verossimilhança pode ser expressa:

$$L(\theta) = \prod_{i=1}^{n} [\lambda(t_i; \theta)]^{\delta i} S(t_i; \theta).$$

2.1.4 Linearização no Modelo de Weibull

Para a seleção do melhor modelo probabilístico podemos lançar mão de uma técnica de seleção gráfica usando as propriedades das função $S(t)$ conforme sua definição abaixo para uma distribuição Weibull:

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

Podendo ser escrito como:

$$\log\{-\log[S(t)]\} = -\gamma \log(\alpha) + \gamma \log(t).$$

Como $\log\{-\log[S(t)]\}$ é função linear de $\log(t)$. Portanto o gráfico de $\log\{-\log[S(t)]\}$ versus $\log(t)$, sendo $\hat{S}(t)$ o estimador Kaplan-Meier, deve haver uma relação linear se o modelo WEIBULL for apropriado.

3 Análise Exploratória dos Dados

Os dados são uma compilação de registros do sistema de Análise de Ocorrências da empresa. Cada unidade geradora é monitorada 24 horas por um centro de operações, que mantêm registros completos de todas as paradas e partidas. As ocorrências de falha com parada da unidade geradora são registrados, e depois analisadas pelas equipes de Operação, Manutenção e Engenharia com objetivo de melhorar as práticas de manutenção e adequar as suas políticas e operação e manutenção.

A base de dados obtida para este estudo é uma planilha eletrônica, com base nestes registros de ocorrências, já previamente analisados, referente ao período do início de operação no ano de 1999 até setembro de 2021 num total de 163 ocorrências com as seguintes variáveis:

1. ID Falha;
2. Número da Unidade Geradora - UG;
3. Data da Falha;
4. Data do Retorno de operação;
5. Equipe de Manutenção da Falha.

Foram fornecidas outras variáveis descritivas dos equipamentos, causa , motivo e tipo desligamento que ficaram de fora do escopo deste artigo.

3.1 Formatação dos dados e exclusões

Foi adotado como premissa que análise seria feita toda na base de tempo em dias, ao se verificar que a base original haviam falhas que duravam menos de 24 horas entre os eventos, foi necessário descartar estes eventos somente são considerados uma única ocorrência de Falha por dia.

Foi necessário converter a informação de falha em cada unidade geradora calculando-se a diferença de tempo entre as datas das falhas $(4.DataRetorno(anterior) - 3.DataFalha(atual))$. Para isto foi usado uma planilha de eletrônica MS Excel onde os dados foram ordenados pela data de ocorrência e foram obtidos uma base com 156 tempos entre a falha atual e anterior sendo a primeira data o inicio de operação comercial da UG.

A variável equipe foi convertida para uma variável indicadora 0 ou 1 da equipe elétrica, que é que têm mais ocorrências e optou-se por fazer uma análise da relação com o tempo até a falha.

O conjunto de dados ajustado agora tem 156 ocorrências com as seguintes variáveis:

1. ID: ID da Ocorrência da Falha;
2. TempoFalha: Tempo de vida (do retorno até ocorrência da falha);
3. UG: Número da Unidade Geradora (UG);
4. Eletrica: Função indicadora (0,1) da Equipe Elétrica;
5. Status: ocorreu a falha =1; como os dados são não censurados todos são 1.

Os dados são **não censurados**, ou seja, todos registros são falhas.

3.2 Resumo de Medidas Descritivas e Gráficos

Para a análise descritiva e análise de sobrevivência foi utilizado o sistema R(R version 4.0.4), e instalados os pacotes adequados (ggplot2, SurvRegCensCov, survival, survminer e patchwork). Foi feita a análise de sobrevivência usando os dados formatados como um dataset usando os seguintes comandos:

```
> # IMPORTAÇÃO  CSV PARA UM DATA FRAME dados1
> require(ggplot2)
> require(SurvRegCensCov)
> require(survival)
> require(survminer)
> require(patchwork)
> dados1 <- read.csv2("hidro_falhas.csv",
+                     header= TRUE,
+                     sep=";",
+                     dec=",")
> dados1$TempoFalha <- as.numeric(dados1$TempoFalha)
> dados1$Status <- as.numeric(dados1$Status)
> str(dados1)
'data.frame':   156 obs. of  5 variables:
 $ ID        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ TempoFalha: num  5 5 3 18 8 29 22 2 37 2 ...
 $ UG        : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Eletrica  : int  1 1 1 0 1 1 1 1 1 1 ...
 $ Status    : num  1 1 1 1 1 1 1 1 1 1 ...
>
> # CÁLCULO DAS MEDIDAS RESUMO
> summary(dados1)
       ID          TempoFalha          UG           Eletrica          Status
 Min.   :  1.00   Min.   :   1.0   Min.   :1.000   Min.   :0.0000   Min.   :1
 1st Qu.: 39.75   1st Qu.:  16.0   1st Qu.:1.000   1st Qu.:1.0000   1st Qu.:1
 Median : 79.50   Median :  64.0   Median :3.000   Median :1.0000   Median :1
 Mean   : 80.46   Mean   : 197.7   Mean   :2.481   Mean   :0.8205   Mean   :1
 3rd Qu.:122.25   3rd Qu.: 217.0   3rd Qu.:4.000   3rd Qu.:1.0000   3rd Qu.:1
 Max.   :162.00   Max.   :1513.0   Max.   :4.000   Max.   :1.0000   Max.   :1
> sd(dados1$TempoFalha)
[1] 317.2808
```

Figura 1: Importação de Dados no R e cálculo medidas

As principais medidas estatísticas resumo para a variável Tempo de vida estão apresentadas na Tabela 1.

Tabela 1: Resumo medidas para a variável Tempo até a falha (dias)

Medida	Valores
Mínimo	1
Máximo	1513
Mediana	64
Média	197,7
Desvio Padrão	317,28
n	156

Podemos analisar graficamente a distribuição da variável tempo de vida com o uso de um histograma e um boxplot, usando os comandos a seguir:

```
> hist1 <- ggplot(dados1, aes(x = TempoFalha)) +
+   geom_histogram(color = "white", fill = "lightblue",bins = 12) +
+   theme_classic(base_size = 18) +
+   xlab("Tempo até a falha (dias)") +
+   ggtitle("Histograma") +
+   ylab("Frequência")
> hist1
> box1 <- ggplot(dados1, aes(x = TempoFalha)) +
+   geom_boxplot(color = "blue", fill = "lightblue" )+
+   theme_classic(base_size = 18) +
+   ggtitle("Boxplot") +
+   xlab("Tempo até a falha (dias)")
> box1
> grafico1 <- box1 + hist1 + plot_layout(nrow = 2)
> grafico1
```

Figura 2: Geração de gráficos Boxplot e Histograma

Nos gráficos apresentados na figura 3, pelo Boxplot vemos que há vários pontos fora do intervalo interquartílicos, e que as diferenças entre a média e mediana se destacam devidos a estes pontos destoantes, mas que são característicos de distribuições assimétricas. Podemos também verificar pelo histograma a distribuição da frequências de ocorrência dos valores da variável TempoFalha entre as classes demonstra uma assimetria a esquerda, mostrando que no intervalo entre 0-500 dias ocorrem grande parte das falhas.

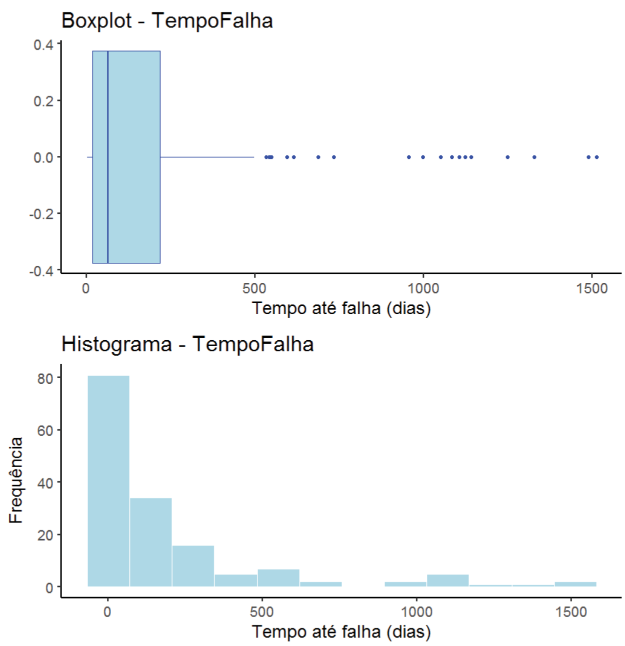

Figura 3: Boxplot e Histograma para a variável TempoFalha.

3.3 Análise Descritiva das Variáveis Auxiliares

Podemos analisar a distribuição da variável **UG** (Número da unidade Geradora) e sua relação com a média da variável **TempoFalha** conforme Tabela 2. Percebemos que a variação é muito pequena entre as UG.

Tabela 2: Resumo medidas para a variável UG x Tempo Médio até a falha (dias)

UG	Qte	%	Média de TempoFalha
1	47	30,1%	172,9
2	28	18,0%	279,8
3	40	25,6%	177,2
4	21	26,3%	190,1
Total	156	100%	197,7

A mesma análise pode ser repetida para a variável **Elétrica** (Indicador Equipe Elétrica , onde 1 = Falha relacionada a equipe elétrica e 0 para as demais equipes) na Tabela 3. A maior parte das falhas teve como responsável a equipe elétrica (82%) e teve grande influência na média geral de 197,7 dias.

Tabela 3: Resumo medidas para a variável Elétrica x Tempo Médio até a falha (dias)

Elétrica	Qte	%	Média de TempoFalha
1	128	82%	218,5
0	28	18%	102,5
Total	156	100%	197,7

As variáveis **Elétrica** e **UG** não se mostraram covariáveis significativas em um modelo de Kaplan-Meier ou Weibull para um p-valor menor que 0,05.

4 Análise de Sobrevivência

Para a análise de sobrevivência foram aplicados os métodos descritos no item 2 - Metodologia.

4.1 Estimador Não paramétrico de Kaplan-Meier

Com o uso do sistema R foi feita a análise de sobrevivência usando os seguintes comandos para obter as estimativas da função de sobrevivência de Kaplan-Meier e o gráfico da Probabilidade de Sobrevivência S(t) x TempoFalha visto na Figura 5:

```
> #  AJUSTE NÃO PARAMÉTRICO DE KAPLAN-MEIER
> ajuste_km <- survfit(Surv(dados1$TempoFalha,
+                           dados1$Status )~1,
+                      conf.type="log",
+                      conf.int=0.95,
+                      type="kaplan-meier",
+                      error="greenwood")
> ajuste_km
Call: survfit(formula = Surv(dados1$TempoFalha, dados1$Status) ~ 1,
    error = "greenwood", conf.type = "log", conf.int = 0.95,
    type = "kaplan-meier")

       n events median 0.95LCL 0.95UCL
[1,] 156    156     64      45     105
> require(survminer)
> p_km <- ggsurvplot(fit = survfit(  Surv(dados1$TempoFalha,
+                                         dados1$Status)~1,
+                                    conf.type="log",
+                                    conf.int=0.95,
+                                    type="kaplan-meier",
+                                    error="greenwood",
+                                    data = dados1),
+                    surv.median.line = "hv",
+                    break.time.by = 100,
+                    xlim = c(0, 1500),
+                    legend = "none",
+                    title = " Função de  Sobrevivência S(t) - Modelo Kaplan-Meier",
+                    xlab = "TempoFalhas(Dias)",
+                    ylab = "Probabilidade de Sobrevivência S(t)")
> p_km
```

Figura 4: Comandos R - para estimativas do modelo não paramétrico de Kaplan-Meier e respectivo gráfico.

A função de sobrevivência (Confiabilidade) obtida pelo estimador não paramétrico de Kaplan-Meier nos fornece a informação que $S(t) = 50\%$ temos um intervalo de confiança de 95% de confiança que o valor do tempo de vida entre 45 e 105 dias.

Graficamente representada a função de sobrevivência $S(t)$ pelo estimador de Kaplan-Meier é apresentada na figura 5. Pode-se verificar seu intervalo de confiança de 95% e sua medina em relação ao tempo até a falha.

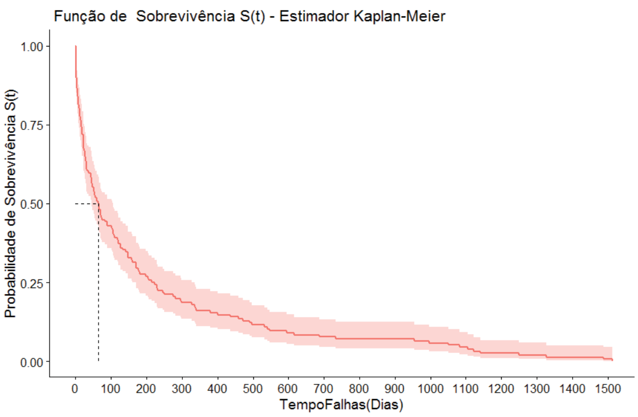

Figura 5: Função S(t) x TempoFalha.

4.2 Modelo Regressão de Weibull

Foram usados os seguintes comandos no Sistema R para a análise de sobrevivência e estimação do modelo de regressão Weibull.

```
> #--------------------------------------------------------------------------
> # AJUSTE DE UMA REGRESSÃO DE WEIBULL
> ajustew<-survreg(Surv(dados1$TempoFalha,dados1$Status) ~ 1, dist='weibull')
> ajustew
Call:
survreg(formula = Surv(dados1$TempoFalha, dados1$Status) ~ 1,
    dist = "weibull")

Coefficients:
(Intercept)
   4.870506

Scale= 1.664657

Loglik(model)= -937.2   Loglik(intercept only)= -937.2
n= 156
> gamma1 = 1/ajustew$scale
> gamma1
[1] 0.6007242
> alpha1 = exp(ajustew$coefficients[1])
> alpha1
(Intercept)
   130.3869
>
>
> #--------------------------------------------------------------------------
> #  PREDIÇÃO USANDO S(T) DO MODELO DE REGRESSÃO PARAMÉTRICO DE WEIBULL
> #  Mediana ou S(t) = 50% sobreviver
> predict(ajustew, type='quantile', p=0.5, newdata = dados1[1,])
      1
70.83745
>
> # CALCULAR O S(t) para 90 dias
> exp(-(90/exp(ajustew$coefficients))^(gamma1))
(Intercept)
  0.4491646
```

Figura 6: Comandos R - para estimativas do modelo paramétrico de regressão de Weibull.

Os parâmetros estimados obtidos para a distribuição de Weibull pelo método da Máxima Verossimilhança em [3] foram $\hat{\alpha} = 130.3869, \hat{\gamma} = 0.6007242$, portanto temos função $S(t)$:

$$\hat{S}(t) = \exp\left\{-\left(\frac{t}{130.3869}\right)^{0.6007242}\right\}.$$

Foi calculado os valores preditos para condições de interesse da empresa que eram:

1-Quantos dias a confiabilidade seria de 50% ;

2-Confiabilidade em 90 dias,

Usando um ajuste de uma regressão para o tempo de falha com a distribuição de Weibull obtivemos os valores estimados para $S(t) = 50\%$ de 71 dias. E para um valor de 90 dias a função $S(t)$ foi de 44,9% .

Com base neste modelo de regressão paramétrico de Weibull podemos construir a curva de Confiabilidade $R(t)$ ou Probabilidade de Sobrevivência $S(t)$ em conjunto com o estimador Kaplan-Meier na Figura 7, e foram usados os comandos no Sistema R a seguir:

```
> #  COMPARAÇÃO GRÁFICA DOS MODELOS KAPLAN-MEIER X WEIBULL
> # GRÁFICO PARA REGRESSÃO DE WEIBULL
> s <- seq(.01, .99, by = .01)
> t <- predict(ajuste_regressao_weibull,
+             type = "quantile",
+             p = s,
+             newdata = dados1[1, ])
> out_wb <- data.frame(time = t,
+                     surv = 1 - s,
+                     upper = NA,
+                     lower = NA,
+                     std.err = NA)
> p_wb <- ggsurvplot(out_wb, conf.int = FALSE, surv.geom = geom_line,
+                   break.time.by = 100,
+                   xlim = c(0, 1500),
+                   legend = "none",
+                   palette = "blue",
+                   title = " Curva de Sobrevivência S(t) x TempoFalhas(Dias)",
+                   xlab = "Tempo Falha(Dias)",
+                   ylab = "Probabilidade de Sobrevivência S(t)")
>
> p_wb
> p_km2 <- ggsurvplot(fit = survfit( Surv(dados1$TempoFalha,dados1$Status)~1,
+                                   type="kaplan-meier",
+                                   data = dados1),
+                    break.time.by = 100,
+                    xlim = c(0, 1500),
+                    conf.int = FALSE,
+                    title = " Probabilidade de Sobrevivência S(t) x TempoFalhas(Dias) , ",
+                    xlab = "TempoFalhas(Dias)",
+                    ylab = "Probabilidade de Sobrevivência S(t)",
+                    legend.title = "Kaplan-Meir (Vermelho) x Weibull(Azul)",
+                    linetype =1)
> grafico_km_x_weibull = p_km2$plot + geom_line(data = out_wb,
+                                              aes(x = time, y = surv),
+                                              color="blue")
> grafico_km_x_weibull
```

Figura 7: Comandos R - para valores preditos do modelo paramétrico de regressão de Weibull.

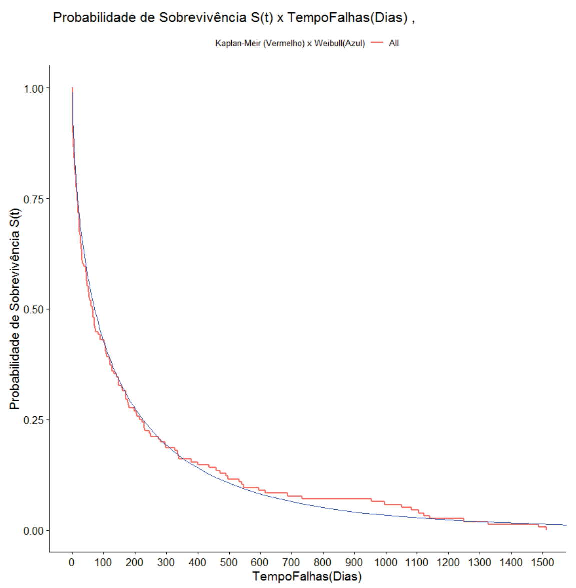

Figura 8: Função Sobrevivência S(t) x TempoFalha.

Pode-se verificar que as duas funções $S(t)$ pelos métodos de Kaplan-Meier e Weibull ficaram bastante próximas, demostrando a viabilidade do modelo Weibull para estimar resultados da confiabilidade dos tempos de falha das UG.

4.3 Adequação do Modelo Weibull

Para a validação do modelo foi gerado um gráfico conforme comandos no Sistema R descritos na figura 9.

```
> # AVALIAÇÃO DO MODELO PROBABILISTICO MÉTODO LINEARIZAÇÃO DE WEIBULL
> require(survival)
> dados<-as.data.frame(dados1)
> i<-order(dados1$TempoFalha)
> dados<-dados[i,]
> ekm<- survfit(Surv(dados$TempoFalha,dados1$Status)~1)
> summary(ekm)
Call: survfit(formula = Surv(dados$TempoFalha, dados1$Status) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    156      12  0.92308 0.02133     0.882195       0.9659
    2    144       4  0.89744 0.02429     0.851068       0.9463
    3    140       2  0.88462 0.02558     0.835875       0.9362

> st<-ekm$surv
> temp<-ekm$time
> invst<-qnorm(st)
> par(mfrow=c(1,1))
> plot(log(temp),
+      log(-log(st)),
+      pch=16,
+      main = "Avaliação Modelo - Linearização Weibull",
+      xlab="log(Tempos)",
+      ylab="log(-log(S(t))")
> yy <- (log(-log(st)))
> yy[112] <- 8
> xx <- log(temp)
> abline(reg = lm(yy ~ xx), col = "red",lty = 1)
>
```

Figura 9: Comandos R - para valores preditos do modelo paramétrico de regressão de Weibull.

A avaliação gráfica da adequação do modelo descrito no item 2.1.4. analisa a relação entre $\log\{-\log[S(t)]\}\, x \log(t)$. Ao verificar os pontos em relação a uma reta x y na figura 10 estas parecem ter uma relação linear, ou seja, o modelo de Weibull parece adequado para este estudo da variável TempoFalha.

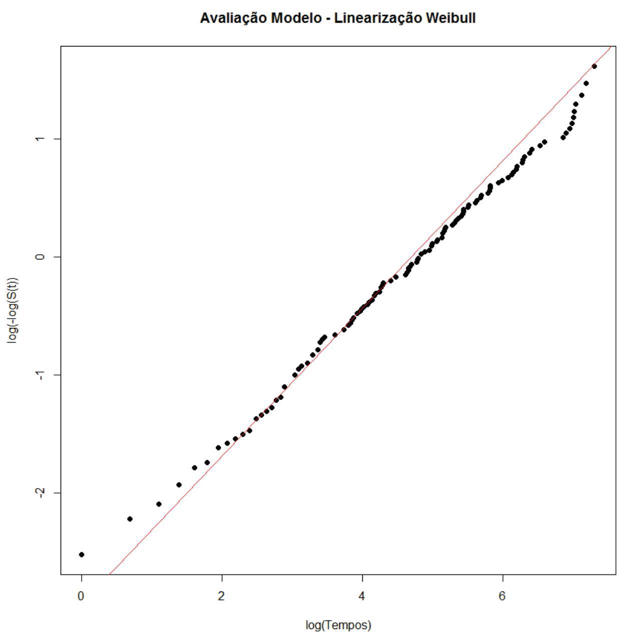

Figura 10: Gráfico da Função $\log\{-\log[S(t)]\}\, x \log(t)$, sendo $S(t)$ o estimador Kaplan-Meier.

5 Conclusões

Os principais resultados obtidos por este estudo foram:

1. Foi construída uma curva da função de sobrevivência (Confiabilidade) versus tempo até a falha. Um

resultado importante é saber a confiabilidade R(90 dias), que é o intervalo médio entre as manutenções preventiva com parada. Para este tempo o modelo de Weibull estimou uma confiabilidade de 44,9% , ou seja, probabilidade acumulada de não ocorrer uma falha do tempo t de 0 até 90 dias é de 44,9% .

2. Aplicação de um modelo probabilístico de análise de sobrevivência que fornece importantes resultados para entendimento do comportamento das falhas e com capacidade de predição.

Agradecimentos

Este artigo apresenta parte dos resultados obtidos com a execução do projeto PD-06491-0341-2014 "Metodologia para gerenciamento de ativos aplicados a hidrogeradores baseado em modelo matemáticos de confiabilidade e manutenibilidade "efetuado pela Universidade Federal Tecnológica do Paraná e a Universidade de São Paulo para a COPEL Geração e Transmissão S.A no escopo do programa de pesquisa e desenvolvimento do Setor Elétrico regulamentado pela Agência Nacional de Energia Elétrica (ANEEL).

Aos colegas e professores da pós graduação de DBSB, em destaque o meu orientador Prof. José Luiz Padilha da Silva.

Faço uma agradecimento especial à minha colega de trabalho e gerente do projeto de PD Gisele Maria de Oliveira Salles pelo apoio e amizade nesta dura caminhada deste projeto.

Referências

[1] EPE Empresa Brasileira de Energia MME Ministério das Minas e Energia. Ben 2022, relatório síntese - ano base 2021. *Site MME*, 2022.

[2] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *Confiabilidade e mantenabilidade - terminologia*. 37p. Rio de Janeiro, 1994.

[3] Flávio Sanson Fogliatto and Ribeiro José Luis Duarte. *Confiabilidade e manutenção industrial*. Elsevier, 2009.

[4] Enrico Antonio Colosimo and Suely Ruiz Giolo. *Análise de sobrevivência aplicada*. Editora Blucher, 2006.

[5] Alexandra Isabel Monteiro Borges. Análise de sobrevivência com o r. *Biblioteca Univesidade Madeira*, 2014.