

# Consultoria Estatística - Grupo 5

Sucateamento de Componentes Eletrônicos



# Sobre a MRS

A MSR é uma operadora logística que administra uma malha ferroviária de 1643 km nos estados de Minas Gerais, Rio de Janeiro e São Paulo, região que concentra metade do PIB brasileiro.

Hoje a companhia está entre as maiores ferrovias de carga do mundo, com produção quase quatro vezes superior àquela registrada nos anos 1990.

Quase 20% de tudo o que o Brasil exporta e um terço de toda a carga transportada por trens no país passam pelos trilhos da MSR.



# Sobre a MRS

A malha ferroviária conecta regiões produtoras de *commodities* minerais e agrícolas e alguns dos principais parques industriais do país aos maiores portos da região Sudeste, o que confere à sua operação uma importância econômica diferenciada.



# Descrição do Problema

Quando os componentes eletrônicos falham, a locomotiva fica avariada, prejudicando o transporte de carga, que é a atividade fim da companhia.

Uma parte dessas falhas têm como origem o fato de itens que já deveriam ter sido sucateados ainda estarem em uso.



**Painel  
TMC**

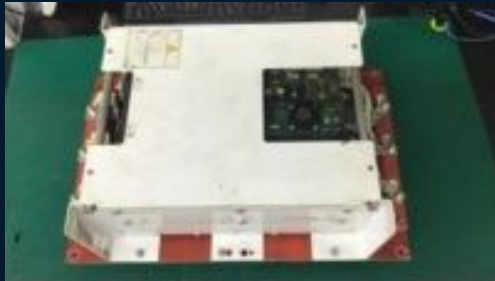


**Painel  
CIO**

# Descrição do Problema

Hoje em dia, não há parâmetros pré-estabelecidos para a decisão de sucateamento ou retirada de circulação dos itens que foram recuperados.

Com isso esse tipo de decisão fica sob a experiência de cada técnico de manutenção, ou seja, não há uma padronização em toda a equipe.



**Painel  
Controlador  
de fase**



**IGBT**

# Problema de Negócio

Criar um sistema/plataforma que auxilie na tomada de decisão sobre sucatear ou retirar de circulação o item, quando inserido o seu número de série (análogo ao CPF) por um colaborador, baseados nos critérios abaixo:

- Identificar a taxa de retorno;
- Identificar as causas do retorno (modo de falha descrito em forma de texto livre);
- Identificar o reparo realizado.



# Dados

Parte do banco de dados vem de formulários digitais (implantados a partir de 2020) e outra parte, gerada manualmente, contendo informações que os técnicos de manutenção mais utilizam.

O banco disponibilizado está filtrado com itens onde consta a informação do seu mês de fabricação.





# Considerações Iniciais



# Colunas do Conjunto de Dados

Abaixo destacamos dentre todas as colunas do conjunto de dados, aquelas que a princípio serão as mais utilizadas:

**Ano** / **DataManut** / LabRespons /Codigo\_Descricao / **Codigo** / Numero Controle  
Lab / Serie / **SerieLimpo** / **DataFabric** / Chave / DiasManut / NumeroOP /  
NumeroOS / Locomotiva / **DescricaoFalha** / DescricaoReparo /  
FichaMaterialRecuperar / FichaMaterialRecuperado / Subcomponente / Falha /  
FalhaLimpo / **Status**



# Colunas do Conjunto de Dados

**Ano:** Ano de registro da manutenção;

**DataManut:** Data que a peça foi avaliada na manutenção;

**SerieLimpo:** Número de Série da peça (código único por peça);

**DataFabric:** Data de fabricação da peça;

**DescricaoFalha:** Breve descrição do problema em texto livre pelo técnico;

**Status:** Status final da peça (Recuperado, Baixado, Outros).



# Dados Faltantes

Um problema inicial na análise se dá pelo fato de que a informação final do estado da peça (**Status**) é faltante em cerca de 46% das observações (1299 valores faltantes), sendo um número muito alto para simplesmente retirarmos do conjunto de dados.

Outro problema se dá ao fato de termos observações de peças que foram Baixadas inúmeras vezes, o que não faz sentido prático.



DataManut	DescricaoFalha	Status
23/7/2020	Realizou inspeção em bancada e testes de tração na locomotiva 3450. Carregou software e verificou as redes de controle e dados.	Baixado
21/8/2020	Ao instalar painel na Jiga constatou falha na comunicação ao mudar a versão do software de PCMCIA para USB e novamente PCMCIA, CPU da rede 0 apresentou defeito causando falha na rede arcnet. Substituiu cartão TMC-1, atualizou e software.	Baixado
25/9/2020	Ao instalar painel na Jiga constatou falha na comunicação ao mudar a versão do software de PCMCIA para USB e novamente PCMCIA, CPU da rede 0 apresentou defeito causando falha na rede arcnet. Substituiu cartão TMC-1, atualizou e software e realizou testes novamente (Inicialização da CPU das redes 0 e1, comunicação nas redes de controle e rede. Verificou configuração).	Baixado

DataManut	DescricaoFalha	Status
30/10/2020	Montou painel na locomotiva 3478 e realizou vários testes: Inicialização nas redes, teste de partida, tração e verificou versão do software. Nada constatado.	Baixado
1/3/2021	Ao verificar a comunicação do painel na jiga constatou que o cartão slot 3 estava com a CPU queimada. Foi substituído o cartão slot 3 por um bom para uso. Testou o painel na loco 3487 e apresentou bom funcionamento. Painel aprovado.	Baixado
13/9/2021	Painel com alta reincidência, será baixado, pois não apresenta boa confiabilidade.	Baixado

# Correção dos Dados Faltantes

Percebemos um certo padrão na coluna de descrição da falha. Os técnicos de manutenção, sempre que baixavam as peças, deixavam clara alguma variação das palavras “Baix” ou “Sucat”, como, por exemplo “Baixada”, “Baixa”, “Sucateamento”.

Usando uma simples expressão regex foi possível classificar uma boa parte das observações, entretanto ainda existem alguns dados ambíguos, uma vez que o texto dos técnicos não condiz com as respectivas datas.



DataManut	DescricaoFalha	Status
28/03/2014	FOI CONSTATADO CPU QUEIMADO E MECANISMO QUEBRADO . DAR BAIXA NÃO TEM PEÇAS DE REPOSIÇÃO .	Baixado
24/03/2015	Foi recuperado mecanismo pcmcia , fez clonagem , realizou testes na giga , fez limpeza do painel parte interna ok .	Recuperado
09/08/2017	RECUPERADO E MODIFICADO MECANISMO PCMCIA, FEITO CLONAGEM DA MEMÓRIA E TESTE NA GIGA OK.	Recuperado

# Correção dos Dados Faltantes

Problemas como os do slide anterior foram resolvidos observando a última data do conjunto de dados. Veja que se a peça foi enviada para manutenção em uma determinada data, é evidente que em todas as suas ocorrências anteriores ela foi recuperada.

Assim, mesmo que a mensagem do técnico seja ambígua, o problema é resolvido e todas as observações passam a ser corretamente classificadas







# Análise de Sobrevivência

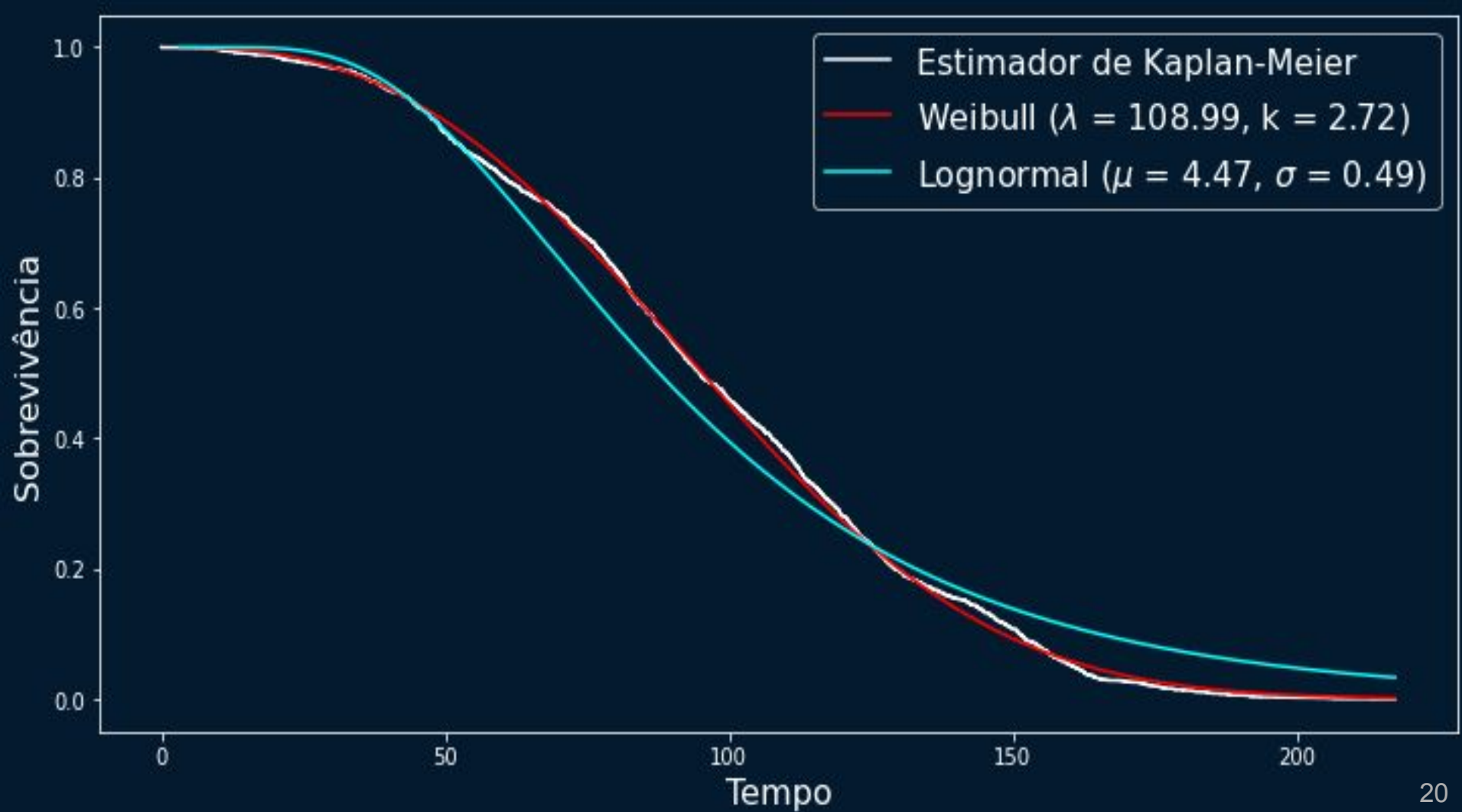
# Análise de Sobrevivência

O passo inicial da análise consistiu em obtermos o tempo da fabricação dos equipamentos até a manutenção. Isso é, subtraímos as colunas **DataManut** e **DataFabric**.

Os dados foram analisado em meses.







Pelo gráfico já fica evidente que o ajuste da distribuição Weibull é superior ao ajuste da Lognormal aos dados. Temos a tabela das estatísticas AIC e BIC:

	<b>AIC</b>	<b>BIC</b>
<b>Weibull</b>	23812.15	23823.68
<b>Lognormal</b>	24379.03	24390.56

Aplicando também o teste de Kolmogorov-Smirnov para os dados e a distribuição Weibull ajustada temos um p-valor de 0.1315, nos levando a não rejeitar a hipótese de que o modelo Weibull é adequado ao nível de significância de 5%.

# Distribuição Weibull

No restante da análise, pretende-se usar como resultado que o tempo de vida entre a manutenção e a fabricação da peça  $T$  é tal que

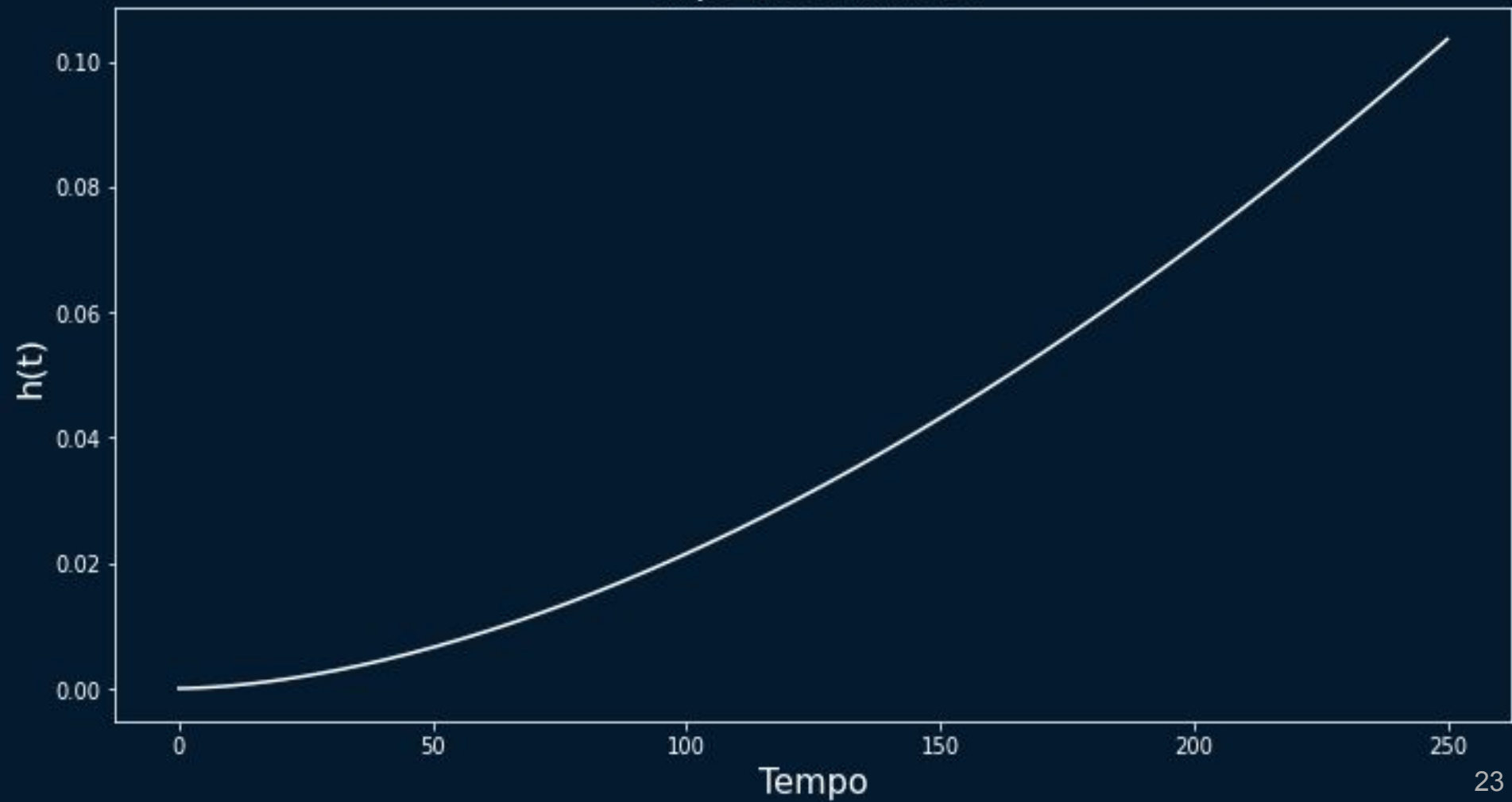
$$T \sim \text{Weibull}(k = 2.7183, \lambda = 108.9899)$$

O que nos permite calcular diversas medidas para cada observação. Medidas que serão úteis na determinação do estado da peça, como por exemplo a função taxa de falha, dada por

$$h(t) = \frac{k}{\lambda^k} t^{k-1}$$



Função Taxa de Falha de T





# Consideração de Diferentes Códigos de Peça



# Diferentes Códigos de Peça

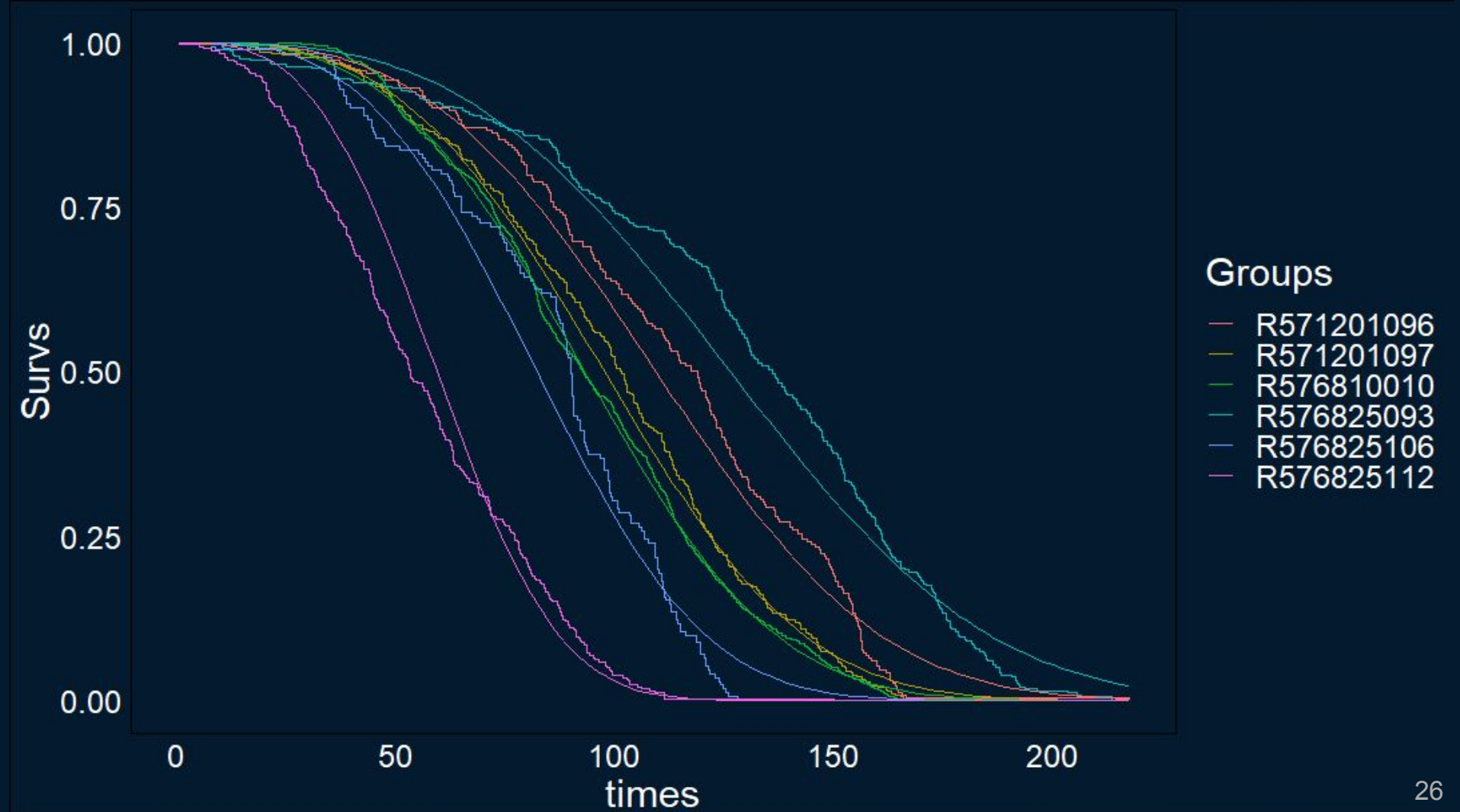
Novamente tomamos a distribuição Weibull, mas dessa vez foi realizado o ajuste de um modelo de teste de falha acelerado

$$S(t) = \exp \left\{ - \left( \frac{t}{\lambda} \right)^k \right\}$$

onde,

$$\lambda = \exp \{ \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \}$$

**Variáveis Dummy  
indicadoras da  
Categoria da Peça**



# Problemas com ambas as abordagens

- Em ambos os casos estamos considerando que cada aparição de uma peça na manutenção é um caso isolado e independente dos demais, mesmo que essa peça já tenha sido levada antes.
- Para uma análise robusta e confiável desses tempos de falha é necessário um método com memória, que leve em conta essas dependências.





# Modelos de Sistemas Reparáveis

# Suposições dos Modelos

- Dado um indivíduo qualquer, consideramos que sua sobrevivência (ou tempo de vida) seja governada por uma distribuição conhecida, com taxa de falha  $\lambda(t)$ . Neste caso, tomaremos a função de um Processo de Lei de Potência (PLP):

$$\lambda(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1}$$

**Mesma função taxa de falha de uma distribuição Weibull( $\beta, \eta$ )**

- Assim que uma falha for observada, é realizado um reparo imediato e instantâneo.
- O reparo pode ser considerado como mínimo, perfeito ou imperfeito.

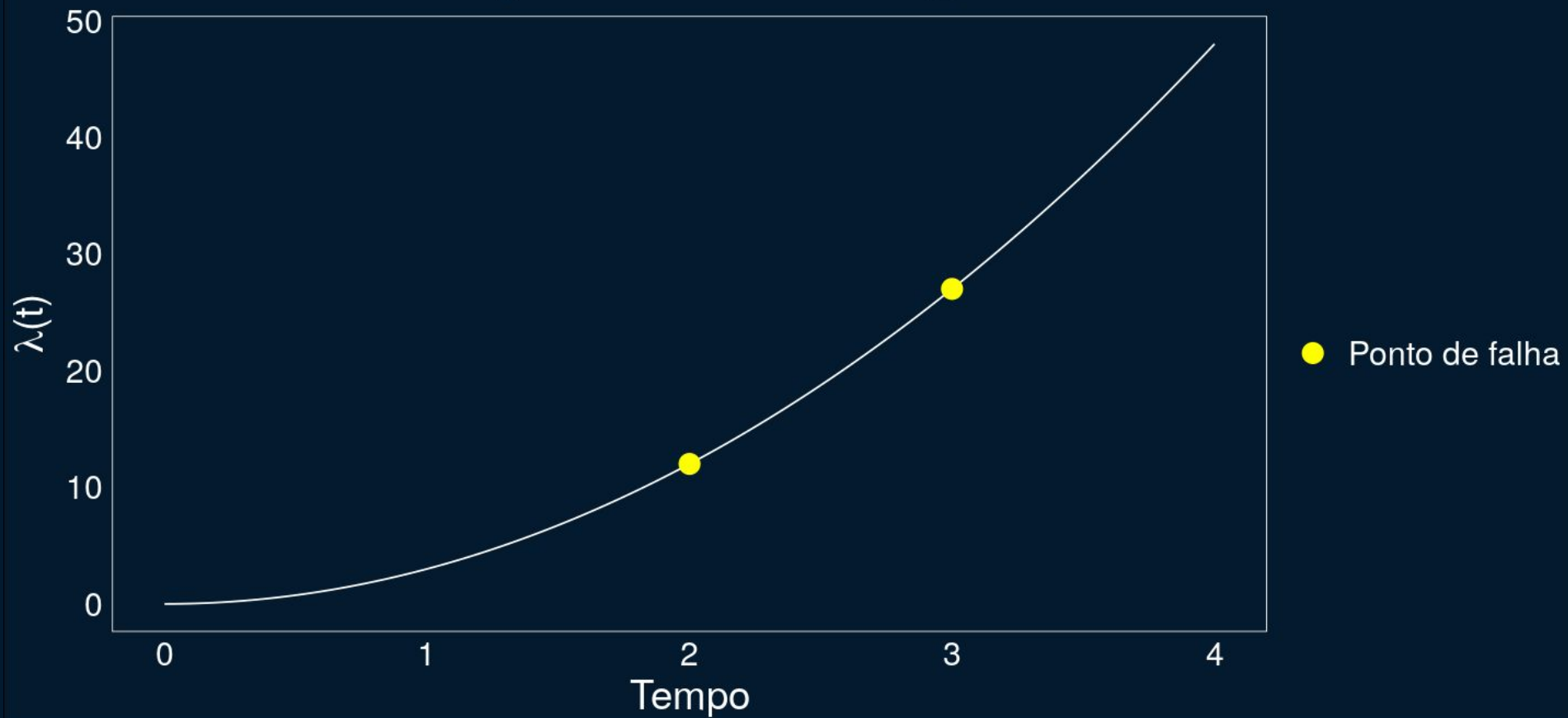


# Reparo Mínimo

- Assume-se que o reparo do objeto é feito de forma mínima. Por exemplo, um computador apresentando falhas é aberto e seu sistema é limpo. Ou uma televisão que está com mau funcionamento e eu dou um tapa para ajustar o sinal.
- Nesse caso, assume-se que não existe alteração da função taxa de falha. Ou seja, o risco de estrago da peça permanece o mesmo de antes após o reparo.



## Função Taxa de Falha $\lambda(t)$



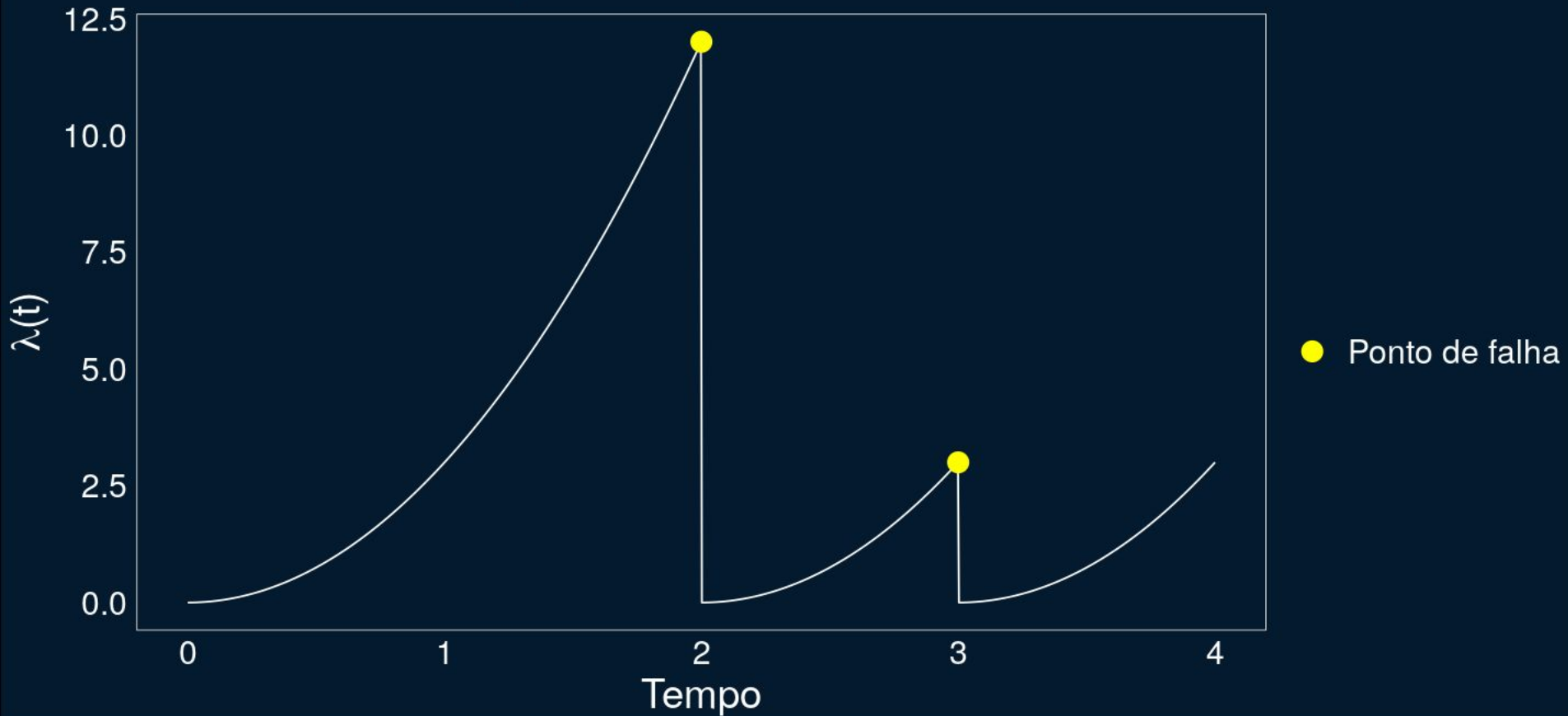
# Reparo Máximo

- Assume-se que foi trocada uma componente importante do equipamento ou quando foi feita uma reforma geral. Por exemplo, quando o celular deu problema, foi levado à loja e teve sua tela e bateria substituídas.
- Nesse caso, assume-se que a taxa de falha volta a se comportar como se o celular estivesse novo em folha.





## Função Taxa de Falha $\lambda(t)$

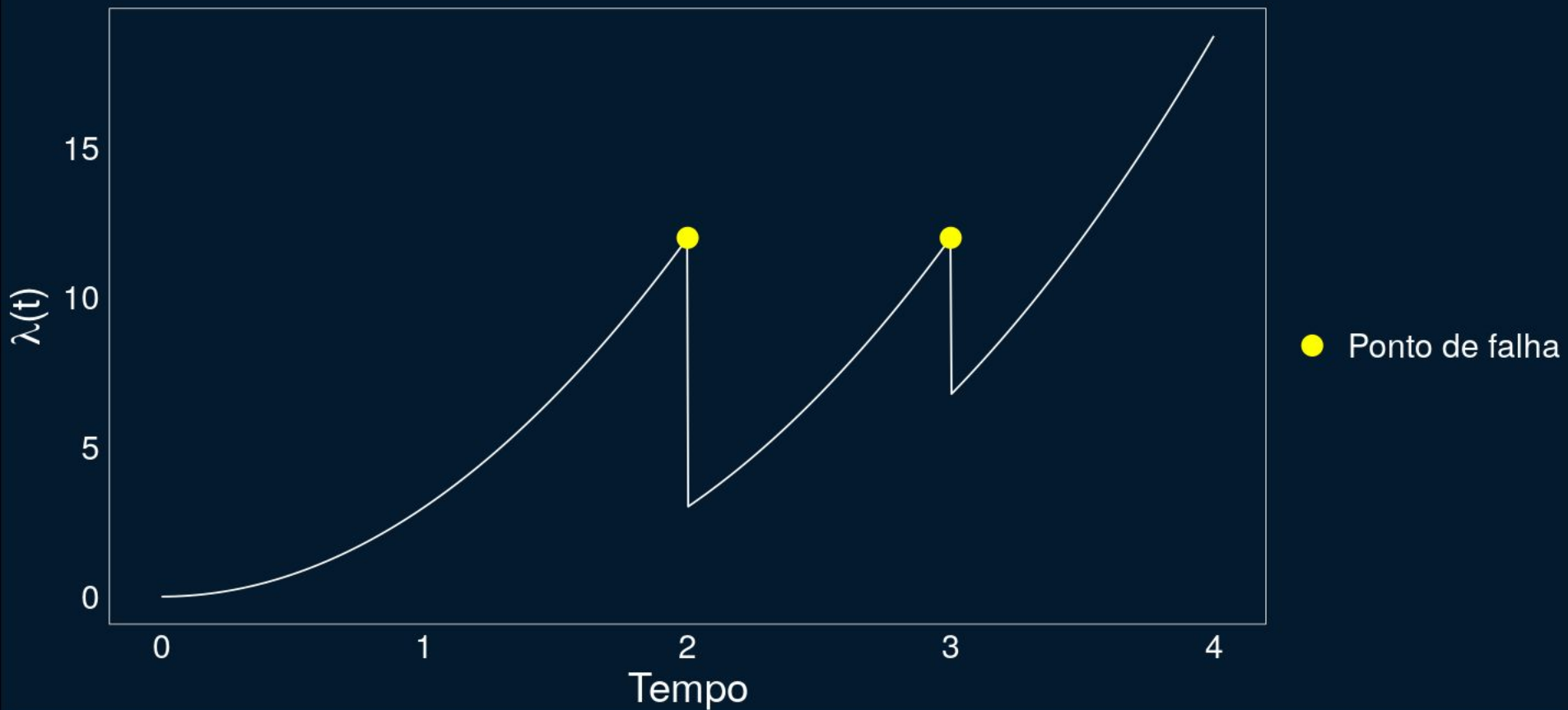


# Reparo Imperfeito

- Assume-se que houve uma melhora do estado geral da peça, mas que também ocorreu um desgaste. Por exemplo, um mouse estava com problemas no clique e no sensor de movimento e foi substituído apenas o sensor de movimento.



## Função Taxa de Falha $\lambda(t)$



# Modelo ARAm

- Existem alguns modelos na literatura que satisfazem as suposições envolvidas e que podem abranger diversos tipos diferentes de reparos. Um desses modelos, que é o utilizado, é o ARAm. Dada a função taxa de falha base  $\lambda(t)$ , a taxa de falha do sistema como um todo será dada por

$$\rho_{ARAm}(t) = \lambda \left( t - (1 - \theta) \sum_{p=0}^{N(t)-1} \theta^p T_{N(t)-p} \right)$$

- Sendo  $N(t)$  o número de falhas já observadas no tempo  $t$  e  $\theta$  um parâmetro flexível.



# Modelo ARAm

- Usando o fato de que, dada a função taxa de falha  $\lambda(t)$ , a função de sobrevivência (ou confiabilidade) de um modelo é dada por

$$\begin{aligned} R_{ARAm}(t_j | t_{j-1}, \dots, t_1) &= \exp \left( - \int_{t_{j-1}}^{t_j} \rho_{ARAm}(u) du \right) \\ &= \exp (\Lambda(a) - \Lambda(b)) \\ &= \exp \left( \left( \frac{a}{\eta} \right)^\beta - \left( \frac{b}{\eta} \right)^\beta \right) \end{aligned}$$

**Lembrando que**

$$H(t) = -\log R(t)$$

$$a = t_{j-1} - (1 - \theta) \sum_{p=0}^{\min(m-1, j-2)} \theta^p t_{j-1-p}$$

$$b = t_j - (1 - \theta) \sum_{p=0}^{\min(m-1, j-2)} \theta^p t_{j-1-p}$$

# Modelo ARAm

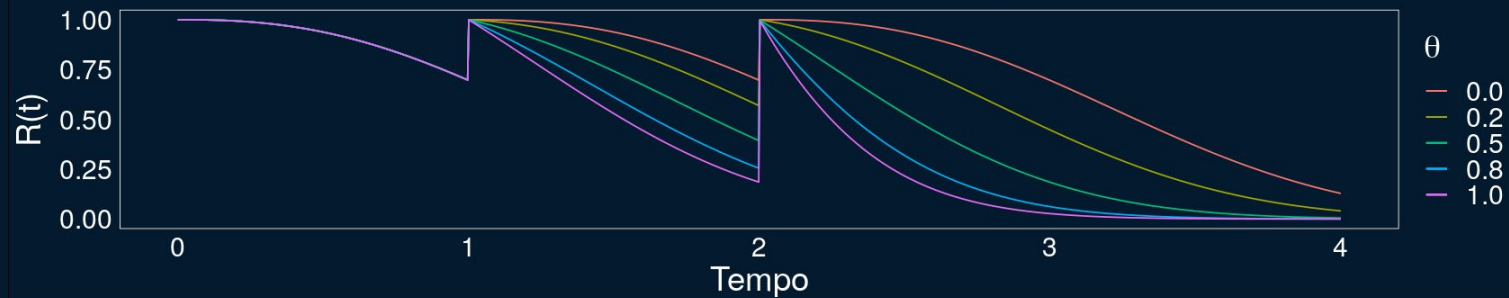
- Um ponto interessante do modelo é que ele toma como casos particulares todos os tipos de reparo.
- Quando  $\theta = 1$ , temos o Reparo Mínimo
- Quando  $\theta = 0$ , temos o Reparo Perfeito
- Quando  $\theta \in (0,1)$ , temos o Reparo Imperfeito



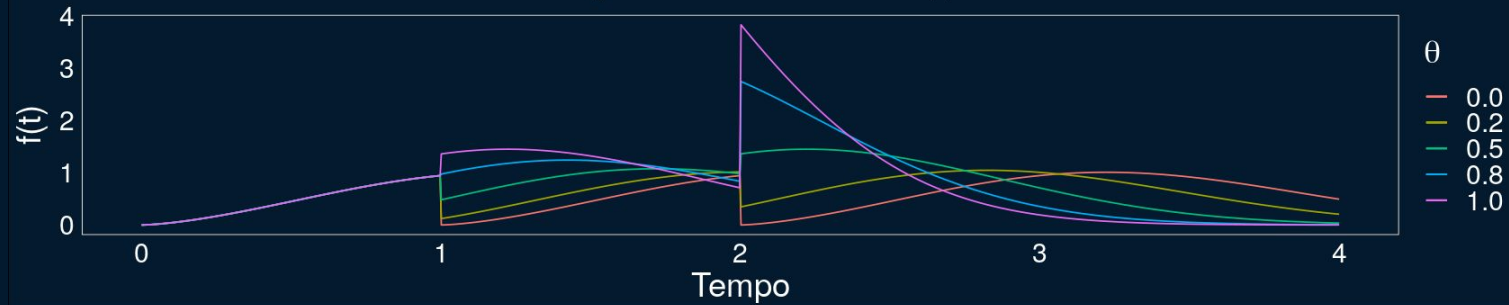
Função Taxa de Falha  $\lambda(t)$



Função de Confiabilidade  $R(t)$



Função de Densidade  $f(t)$





# Construção da Verossimilhança





# Função Log da Verossimilhança

A contribuição da peça  $i$  para a verossimilhança se dará por

$$f(t_{i,1}, \dots, t_{i,n_i}) = f(t_{i,n_i}|t_{i,n_i-1}, \dots, t_{i,1}) \cdots f(t_{i,j}|t_{i,j-1}, \dots, t_{i,1}) \cdots f(t_{i,2}|t_{i,1})f(t_{i,1})$$

Em que  $n_i$  representa o número de falhas da peça  $i$  e os subíndices representam os tempos em que ocorreram as falhas da peça.

$$L(\beta, \eta, \theta|m, \mathbf{T}) = \prod_{i=1}^N \prod_{j=1}^{n_i} f(t_{i,j}|t_{i,j-1}, \dots, t_{i,1})$$





# Amostragem do Modelo

# Amostragem do Modelo ARAm

A partir da função de confiabilidade apresentada anteriormente

$$R(t_j | t_{j-1} \dots t_1) = \exp(-\Lambda(a) + \Lambda(b))$$

Podemos obter a função densidade acumulada e a inverter, obtendo assim uma expressão que permite a amostragem do modelo pelo método da inversa.

$$F^{-1}(z) = \Lambda^{-1}(\log(1 - z) + \Lambda(a)) + (1 - \theta) \sum_{p=0}^{\min\{m-1, j-2\}} \theta^p t_{j-1-p}$$



# Amostragem do Modelo ARAm

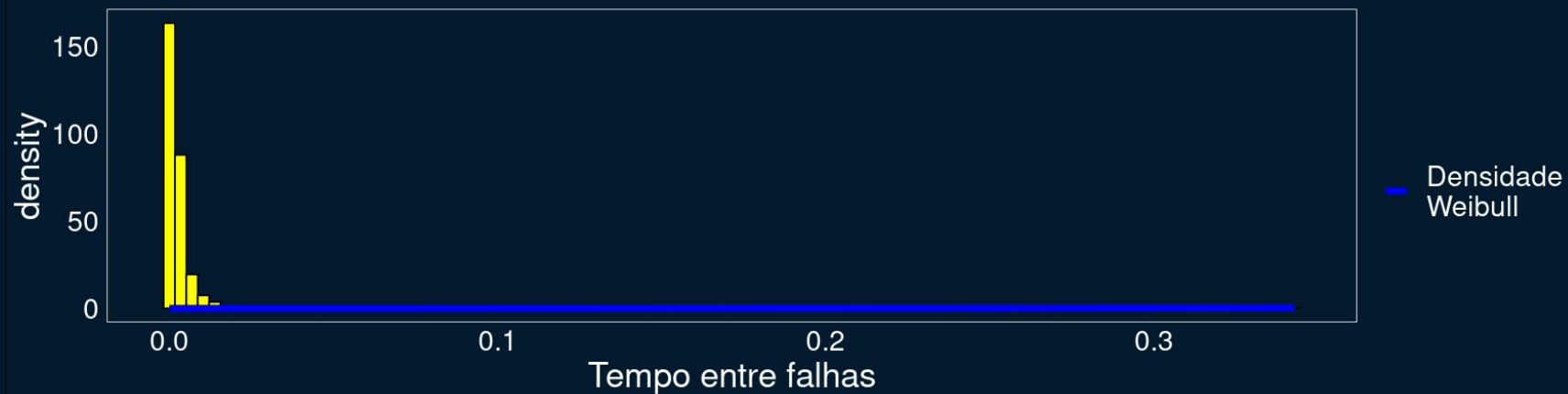
Sendo a distribuição inicial caracterizada por uma Weibull, segue que após amostrado o primeiro tempo de falha atualizamos a distribuição para o próximo. Esse processo pode ser repetido o número de vezes que for desejado.

Amostra 1 Peça	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
$\theta = 0$	0.8321	2.4149	3.1328	3.5954	4.5714	5.3045	6.2645	6.9556	8.6908	9.7363
$\theta = 1$	1.3929	1.5512	1.6175	1.6732	1.7003	1.8387	1.9031	1.9599	2.0371	2.0857

$\theta = 0$



$\theta = 1$





# Estudo de Simulação

# Estimação

Sendo bem definida a função de verossimilhança, basta definirmos uma função numérica para a estimação dos parâmetros  $\beta$ ,  $\eta$  e  $\theta$  maximizam o valor do logaritmo dessa função com base na amostra dos dados.

Para verificarmos o ajuste dos modelos, fazemos um rápido estudo de simulação dos dados, verificando se o valor estimado de um conjunto de dados gerado aleatoriamente converge para os valores reais pré definidos do modelo



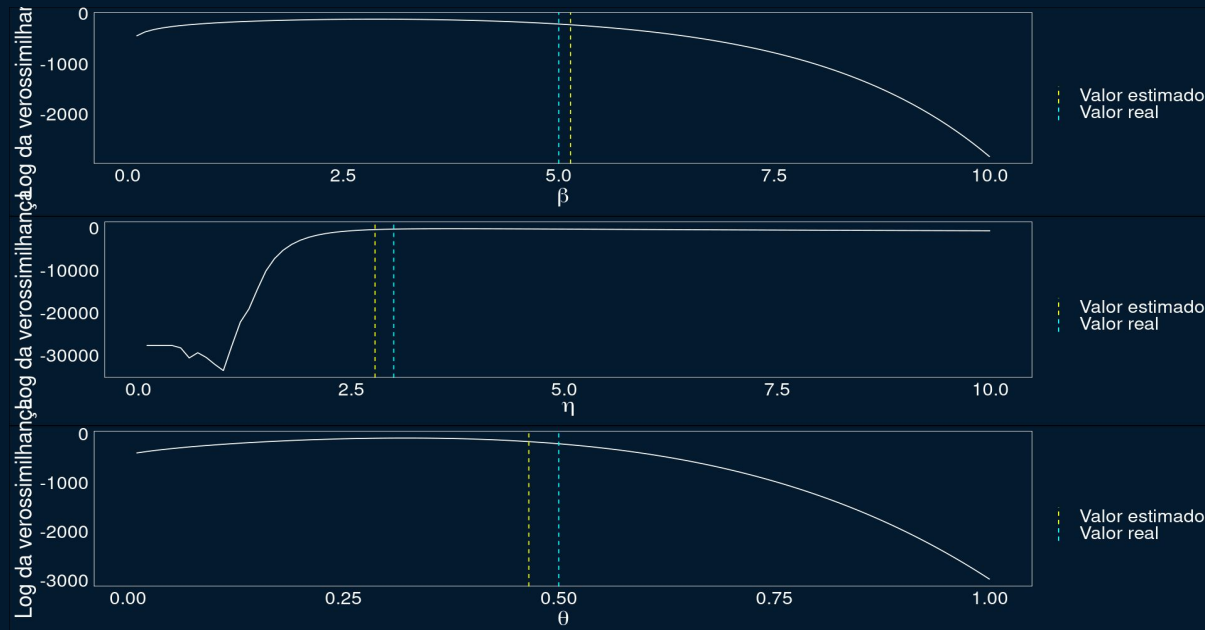


# Simulação

Consideramos  $\beta = 5$ ,  $\eta = 3$ ,  $\Theta = 0.5$ . Gerando com esses parâmetros uma amostra de 20 peças, cada uma com 6 falhas. Consideremos um modelo ARAm com  $m = 2$ .

Peça	T1	T2	T3	T4	T5	T6
1	3.4076	3.9846	5.6708	6.5794	7.9324	8.6862
2	2.0774	3.4006	4.5691	5.5524	6.6752	7.3907
3	2.7121	3.9066	5.3615	5.8543	7.7469	9.0992
...	...	...	...	...	...	...

Parâmetros	$\beta$	$\eta$	$\Theta$
Valores Reais	5	3	0.5
Valores Estimados	5.1365	2.7803	0.4653





# Exemplo de Aplicação

# Manutenção de Caminhões

Conjunto que contabiliza o reparo de 5 caminhões individuais ao longo do tempo, no total os caminhões passaram por algum tipo de reparo de 23 a 32 vezes antes de serem baixados (ou antes da coleta de dados ter se encerrado). Um exemplo dos tempos de falha e reparo do caminhão 1:

9.197, 14.761, 19.246 24.597 25.302 30.948 31.602 34.574 41.059  
44.114 44.279 46.381 54.519 56.355 74.44 77.236 80.261 82.815  
84.772 88.901 93.727 98.149 106.429



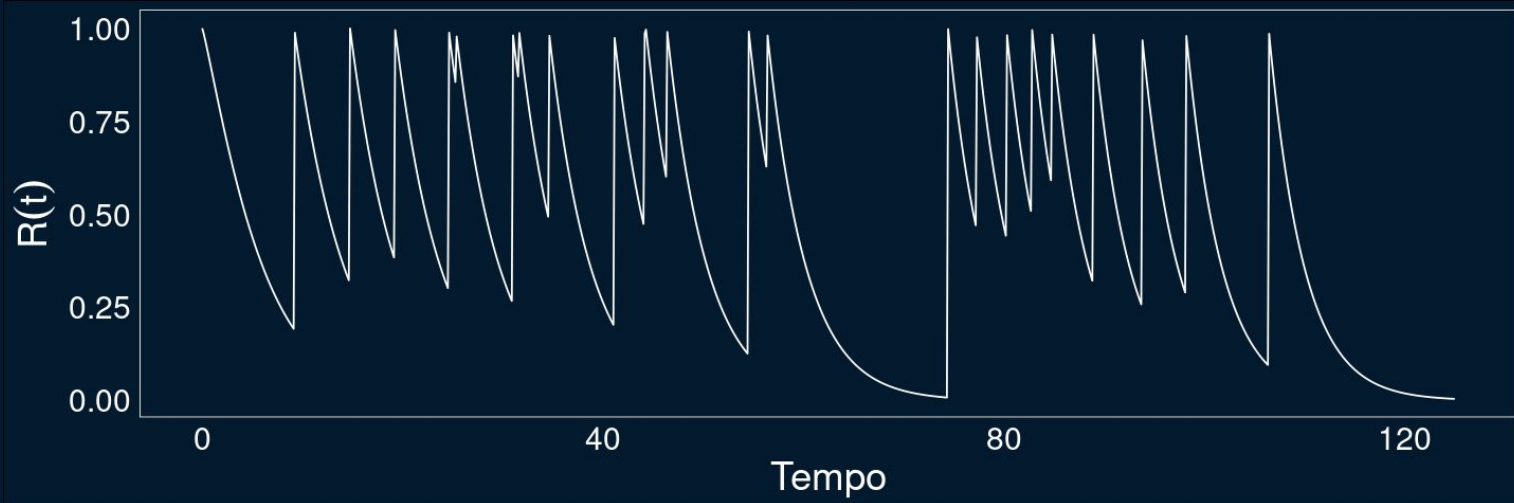
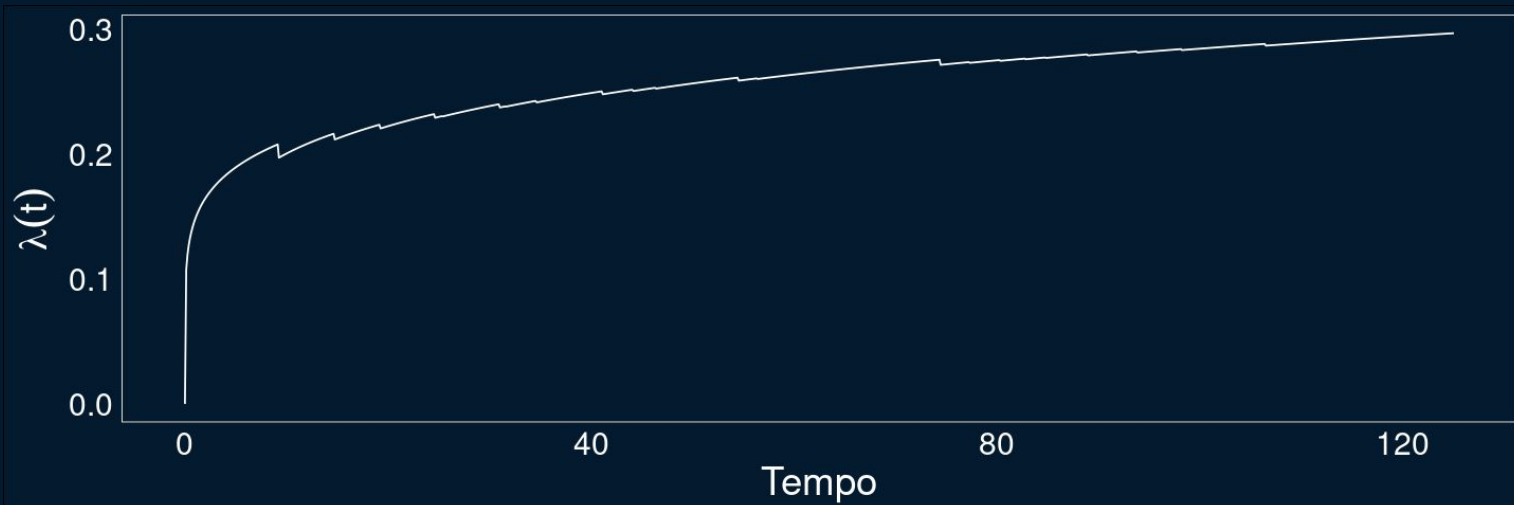
# Manutenção de Caminhões

Quando aplicados os dados dos caminhões no modelo ARA1 temos

Parâmetros	$\beta$	$\eta$	$\Theta$
Valores Estimados	1.1538	5.929	0.6991

Nesse caso vemos que  $\Theta$  indica que o tipo de reparo dos caminhões se situa mais próximo do reparo mínimo do que do reparo perfeito. Vejamos o gráfico da curva de risco e de sobrevivência estimadas para o primeiro caminhão







# Aplicação aos dados da MRS

# Aplicação

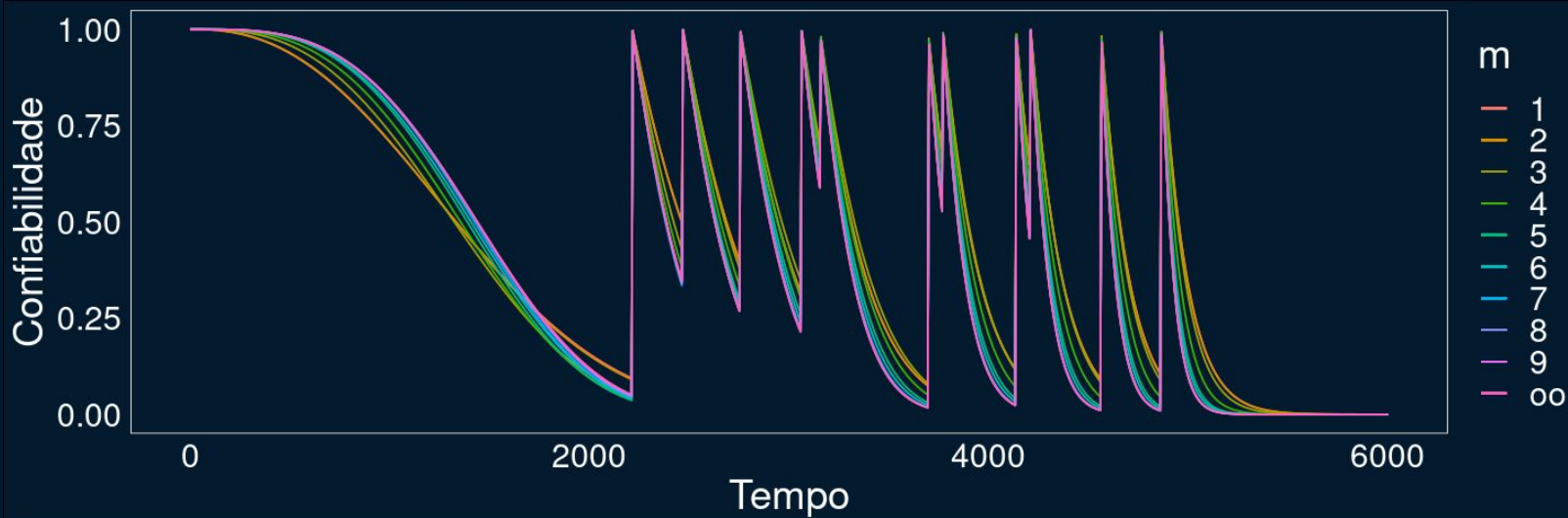
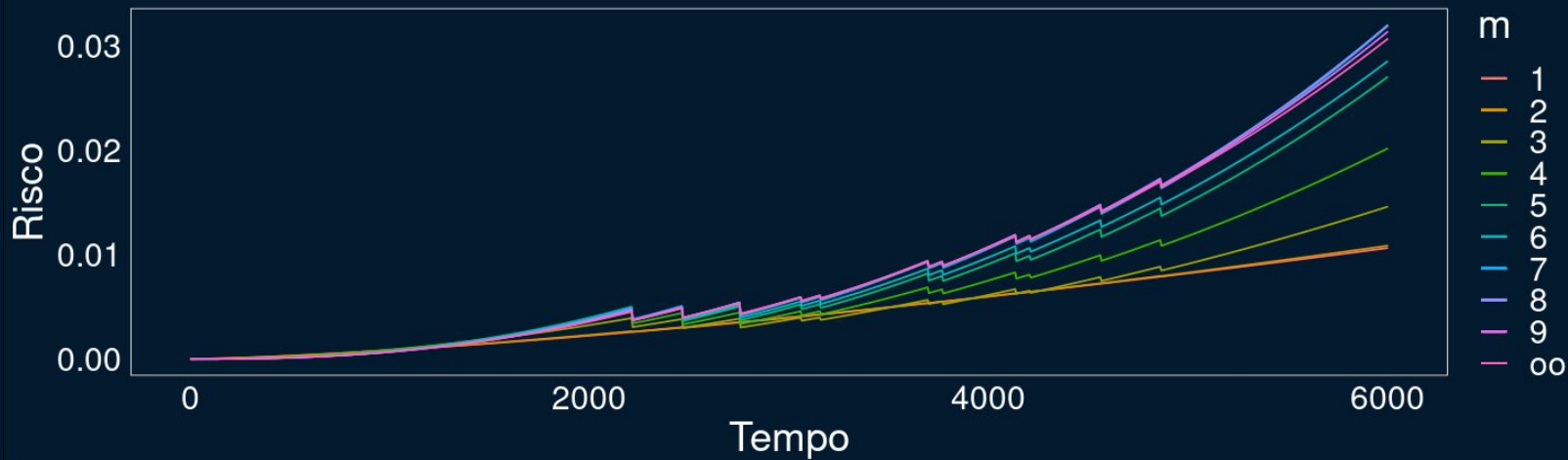
Primeiramente vejamos o comportamento do modelo para uma pequena quantidade de observações. Foram tomadas apenas as quatro peças que observaram o maior número de falhas. Todas essas peças pertencem a um mesmo tipo (Código R576810010).



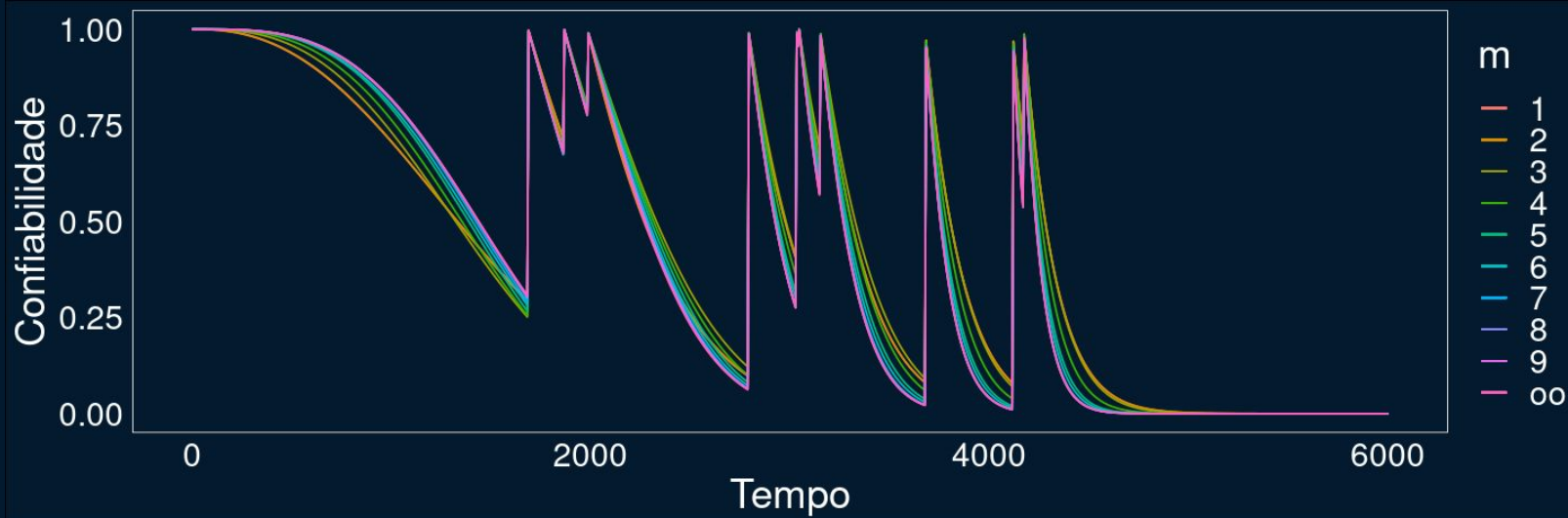


Parâmetros	$\beta$	$\eta$	$\Theta$
$ARA_1$	2.4173	1549.5897	1
$ARA_2$	2.446	1544.6477	0.9833
$ARA_4$	3.1215	1522.0736	0.8731
$ARA_6$	3.3658	1567.9833	0.8925
$ARA_8$	3.4386	1596.1403	0.9056
$ARA_\infty$	3.3825	1602.2439	0.9144

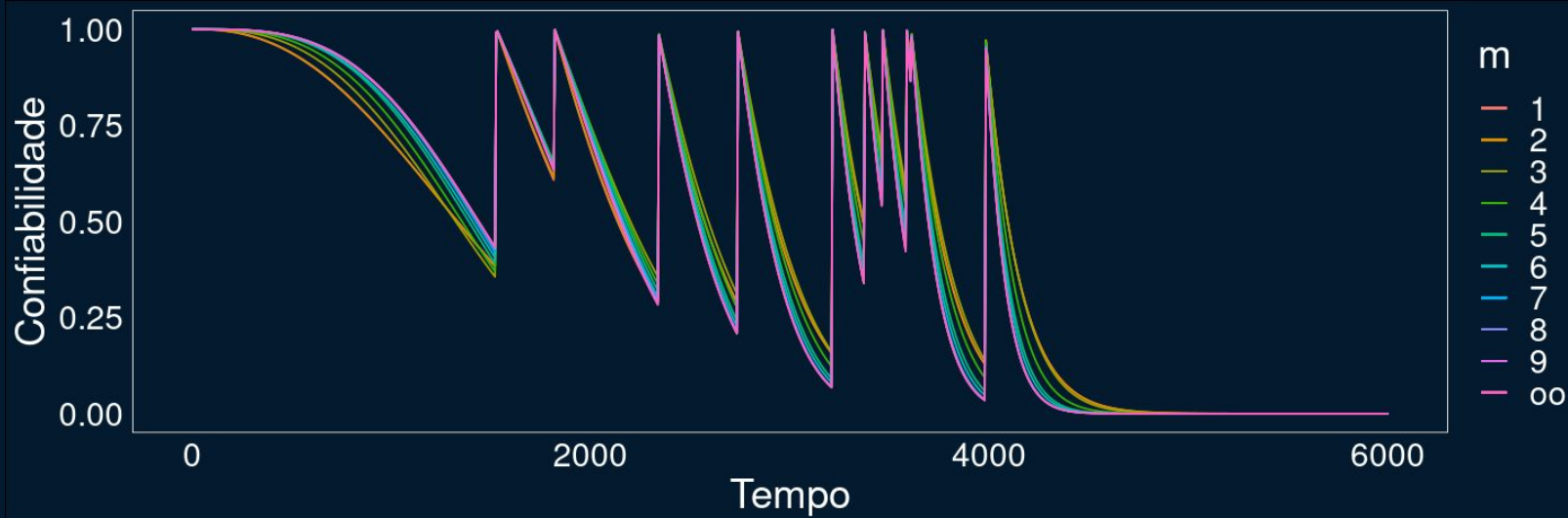
# FM08080018



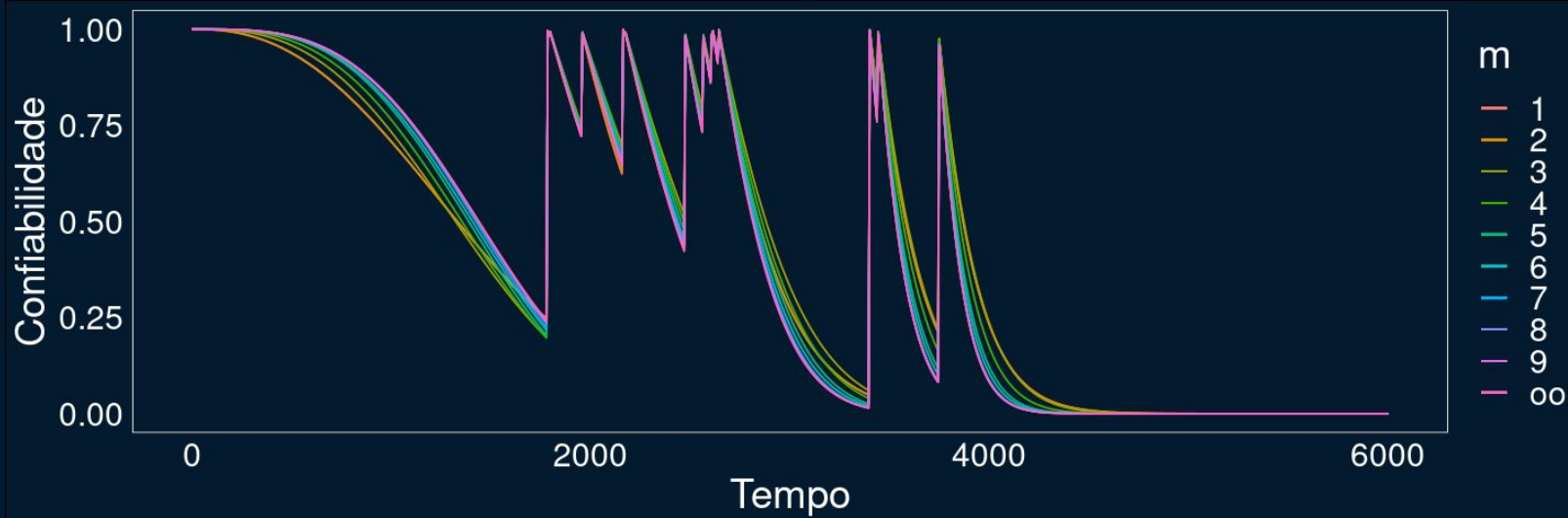
FM10074893



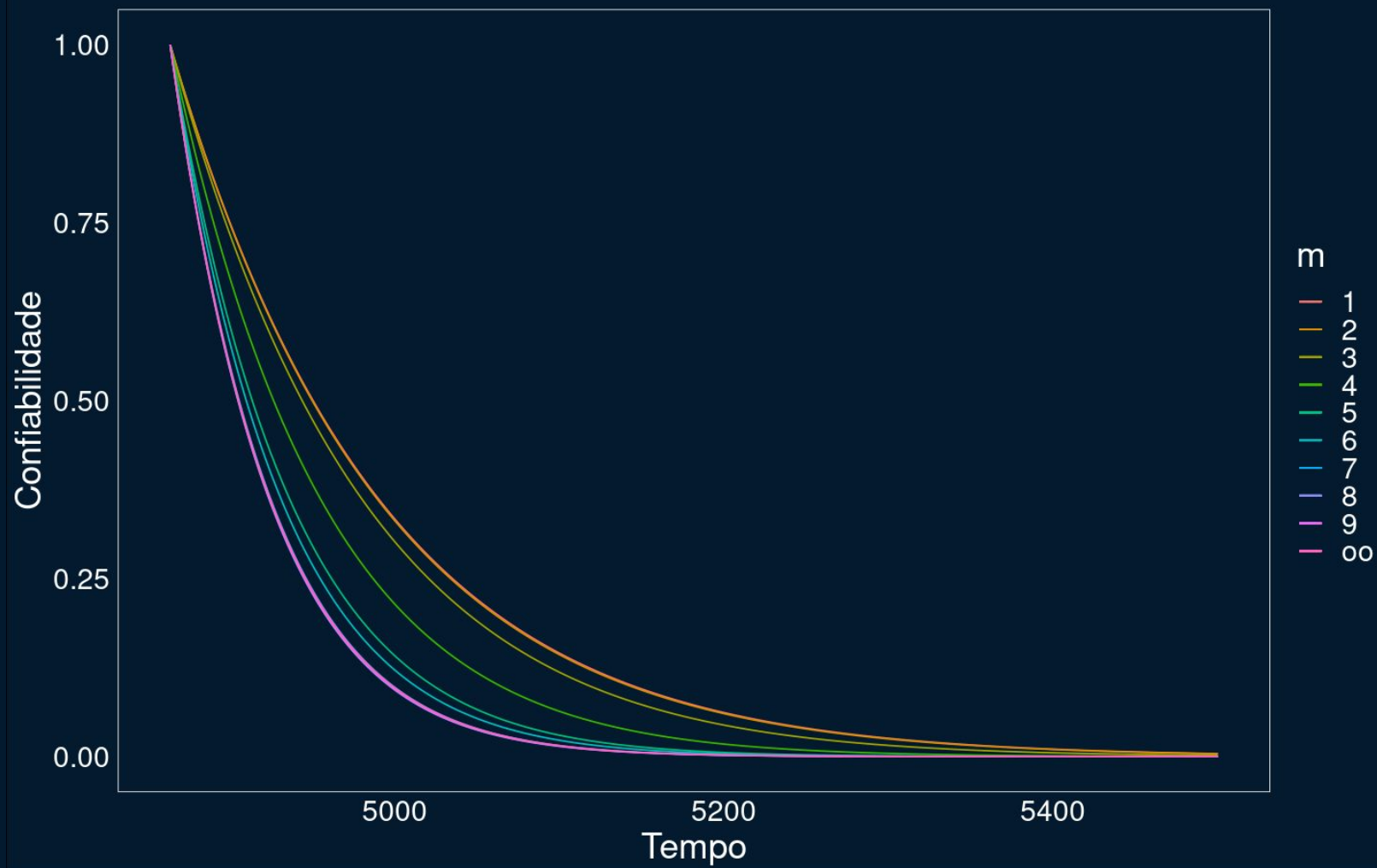
FM11024833



FM11044661

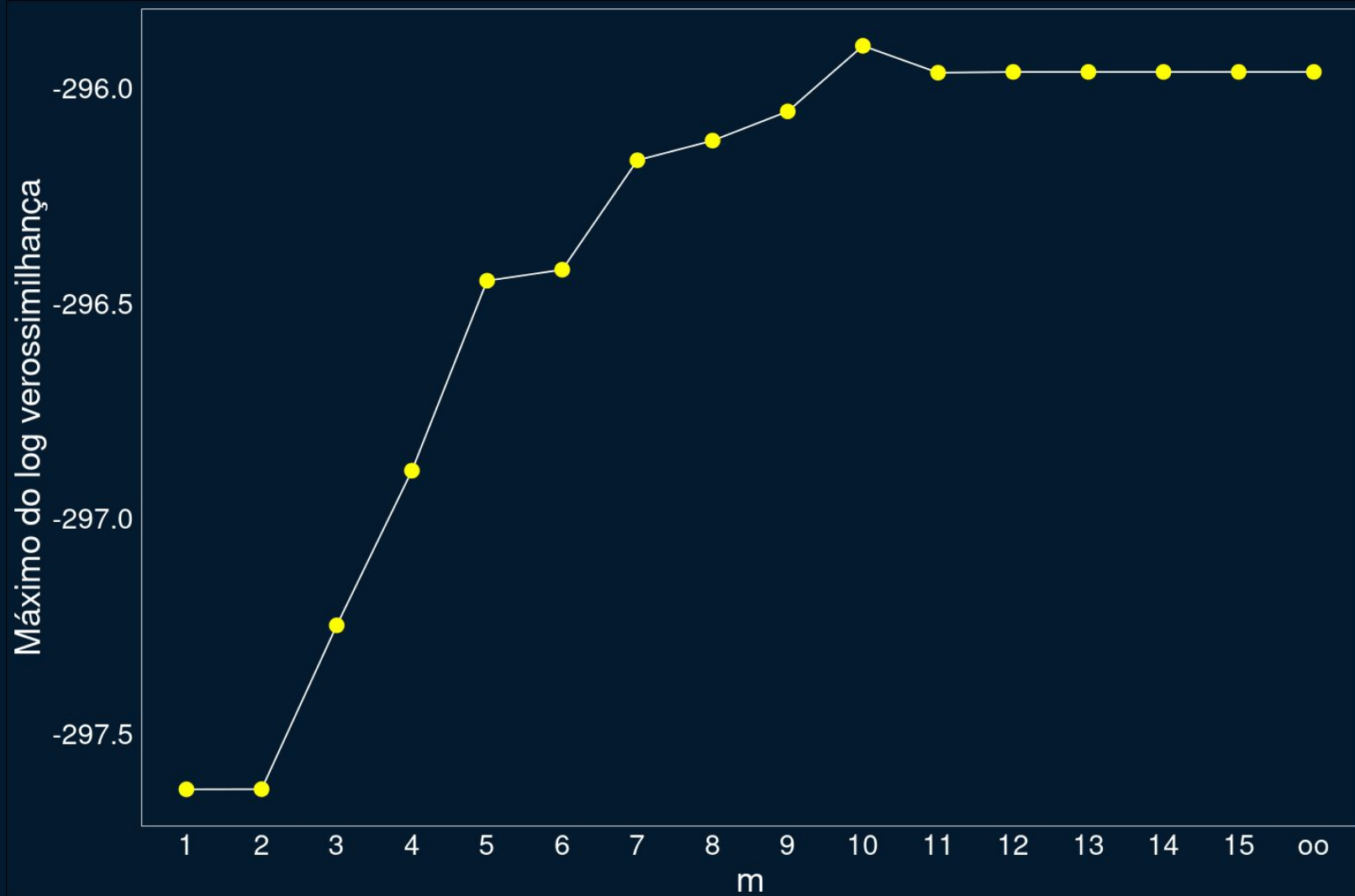


FM08080018



# Escolha do valor $m$

Ajustando o modelo para  $m = 1, 2, \dots$  é esperado que após um determinado valor os modelos se estabilizem, convergindo para um ponto específico. Para os pontos ajustados







# Formatação Final Tabela de Dados

# Aplicação

Para a aplicação final do modelo aos dados, separamos as peças em 6 grupos, sendo cada um composto segundo um tipo de peça.

Um último ponto a ser levado em consideração deve ser a presença de observações censuradas: Suponha que uma peça falhou uma única vez, em 2020 e nunca mais retornou para manutenção. No conjunto de dados temos apenas a informação de falha dessa peça.



# Censuras

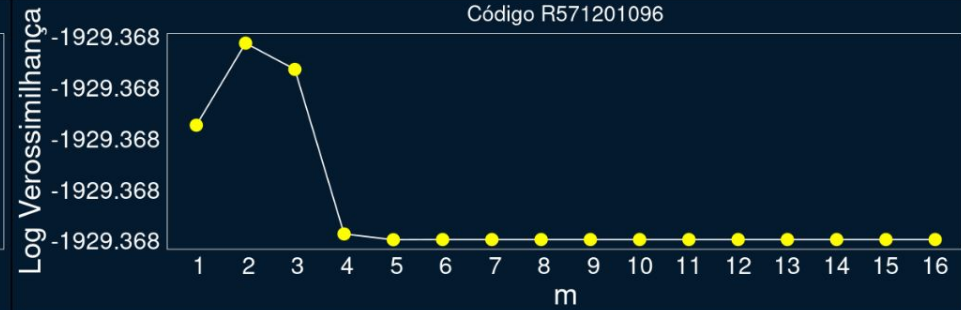
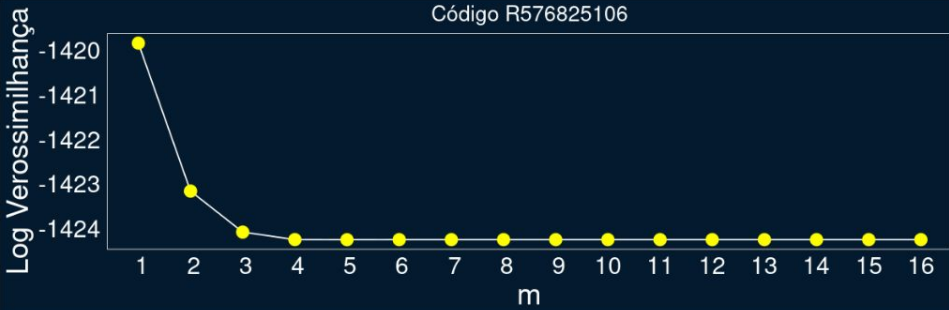
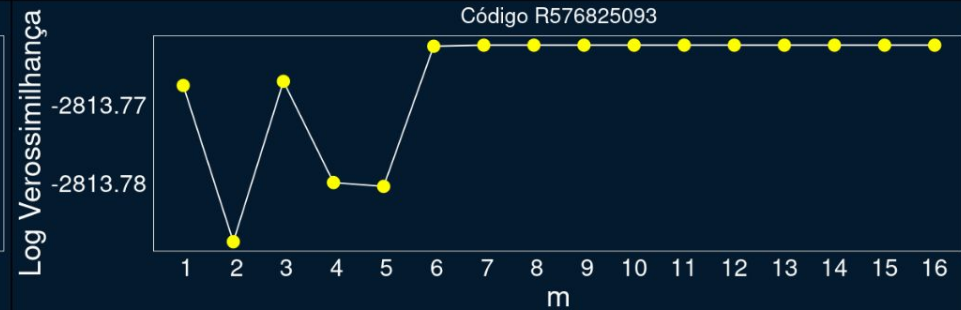
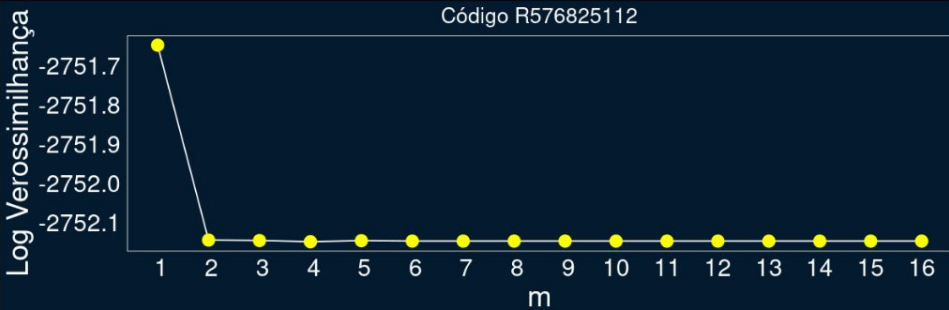
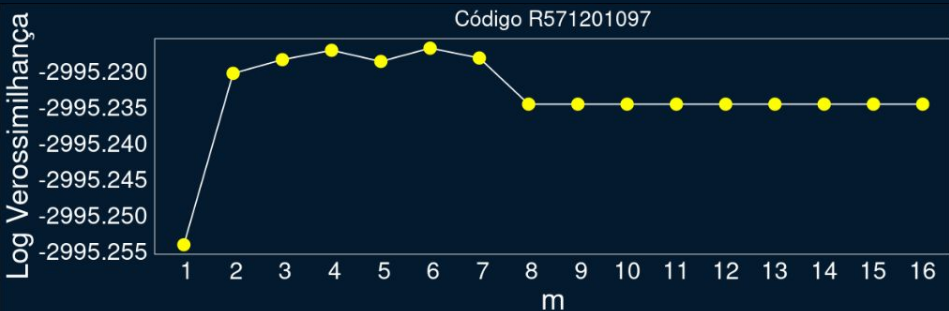
A função de verossimilhança que passou a ser considerada é dada por

$$L(\beta, \eta, \theta | m, \mathbf{T}) = \prod_{i=1}^N \prod_{j=1}^{n_i} f(t_{i,j} | t_{i,j-1}, \dots, t_{i,1})^{\delta} R(t_{i,j} | t_{i,j-1}, \dots, t_{i,1})^{1-\delta}$$

onde  $\delta=1$  representa que houve uma falha da peça no tempo  $t_{i,j}$  e  $\delta=0$  indica que até o momento não houve nenhuma falha. Essa informação é muito útil para o modelo, podendo trazer interpretações completamente diferentes.



Índice	Nome da Peça	Tempo	Tempos Anteriores	$\delta$
1	FL08080101	342	0	1
1	FL08080101	4947	342	0
2	FL10064357	1365	0	1
2	FL10064357	1596	1365	1
2	FL10064357	1603	1365, 1596	1
2	FL10064357	3125	1365, 1596, 1603	1
2	FL10064357	3678	1365, 1596, 1603, 3125	1
2	FL10064357	4278	1365, 1596, 1603, 3125, 3678	0
...	...	...	...	...



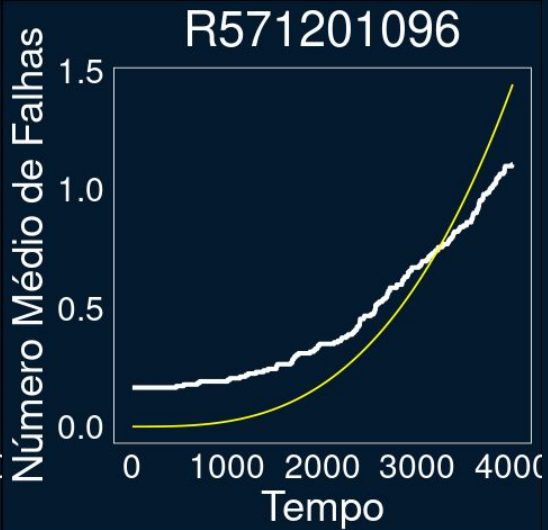
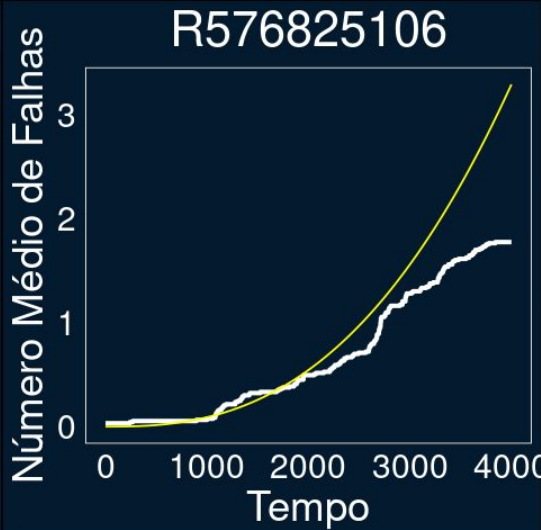
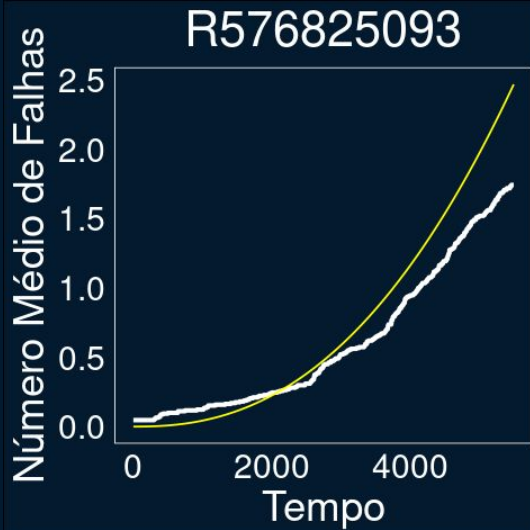
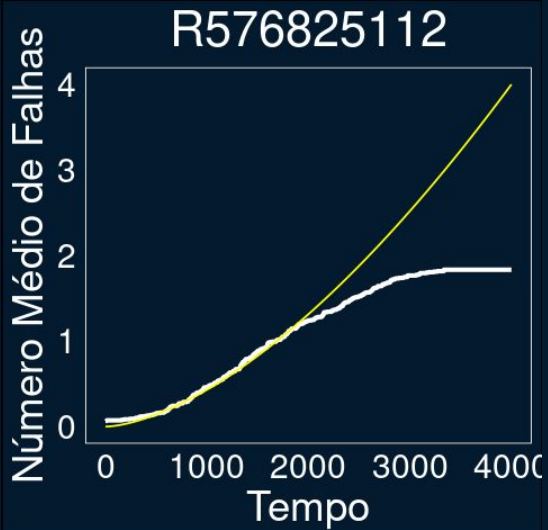
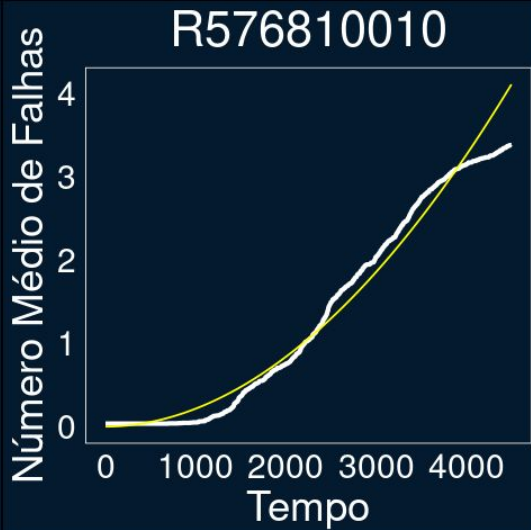
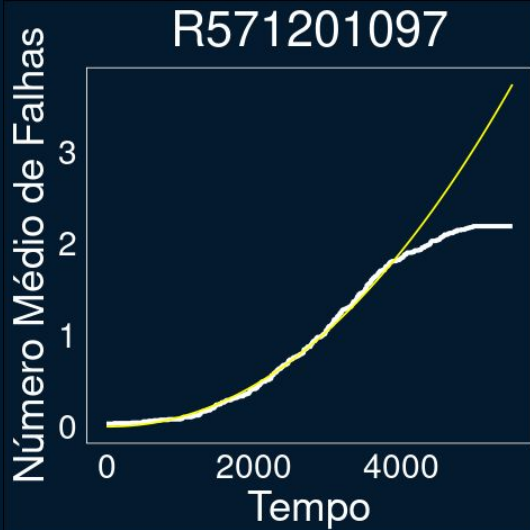
Código	m	$\beta$	EP( $\beta$ )	$\eta$	EP( $\eta$ )	$\theta$	EP( $\theta$ )
R571201097	6	2.0921	0.1038	2931.493	94.8098	1.0000	0.0017
R576810010	2	1.9596	0.0524	2186.388	51.9897	1.0000	0.0003
R576825112	1	1.6057	0.0871	1685.568	77.0812	0.7886	0.1768
R576825093	6	2.3725	0.1207	3758.135	108.7710	1.0000	0.0005
R576825106	1	2.6000	0.2016	2529.857	100.3434	0.5435	0.0720
R571201096	2	3.0215	0.1733	3543.094	87.8251	1.0000	0.0000

# Diagnóstico do Ajuste

Para verificarmos se o ajuste do modelo aos dados pode ser considerado aceitável construímos um gráfico comumente usado na literatura.

- Para cada tempo  $t$ , obtém-se um número médio de falhas das peças que ocorreram até  $t$  (curva empírica dos dados)
- Construção da curva da média da integral do risco da última falha até o tempo  $t$ ; Em boa parte dos modelos ( $\theta = 1$ : Reparo Mínimo) podemos substituir esse cálculo simplesmente pelo gráfico do risco acumulado sem falhas anteriores
- O risco acumulado procura neste caso prever o número de falhas das peças







# Diagnóstico do Ajuste

- A partir dos gráficos identificamos um problema no ajuste do modelo: o número de falhas predito ao longo do tempo ultrapassa o número real de falhas
- A principal hipótese para a causa desse problema é o fato de termos muitas peças que apresentaram apenas uma ou poucas falhas no conjunto de dados
- Uma vez que a teoria se estabiliza para grandes tamanhos amostrais de cada peça, esse fato contribui negativamente para o ajuste



# Diagnóstico do Ajuste

- Embora tenhamos esses problemas, veja que inicialmente a curva predita pelo modelo é muito próxima à curva empírica, indicando que o modelo é capaz de prever relativamente bem a distribuição das peças em tempos relativamente baixos. Quanto maior o tempo, mais incerta é a previsão





# Apresentação da Análise

# Apresentação da Análise

- Para a apresentação dos resultados de forma comercial e direta faremos o uso da plataforma PowerBI, construindo os diagnósticos para cada peça e fornecendo algumas métricas interessantes como o tempo médio de vida estimado para a peça a partir do momento do último reparo, por exemplo.
- Ao final temos uma aplicação para a consulta de peças na base de dados em conjunto com um relatório resumido das principais informações dessas peças, as quais poderão ser úteis para os técnicos tomarem suas decisões durante o reparo.





# Planejamento Semanal

	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10
<b>Seleção de Variáveis</b>										
<b>Análise Gráfica</b>										
<b>Identificação da Distribuição dos Dados</b>										
<b>Coluna Status</b>										
<b>Separação dos Tipos de peça</b>										
<b>Análise de Recorrência</b>										
<b>Criação das Métricas quantitativas</b>										
<b>Modelo de Classificação (Status)</b>										
<b>Desenvolvimento do Modelo do Power BI</b>										
<b>Versão Final</b>										