



Instituto Superior de Engenharia

Politécnico de Coimbra

DEPARTAMENTO DE ENGENHARIA QUÍMICA E
BIOLÓGICA

Manutenção Preditiva de uma Prensa de Secagem

Dissertação para a obtenção do grau de Mestre em Engenharia e
Gestão Industrial

Autor

Francisco Carlos Guedes Paiva

Orientador

José Manuel Torres Farinha

Co-Orientador

Luís Manuel dos Santos de Melo Margalho



INSTITUTO POLITÉCNICO
DE COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Coimbra, dezembro de 2023

AGRADECIMENTOS

Este trabalho não teria sido possível sem o apoio e orientação do Professor Doutor José Torres Farinha, bem como do Professor Doutor Luís Melo Margalho. Agradeço o contributo e acompanhamento ao longo de todo o desenvolvimento desta dissertação.

A todos os colegas, familiares e amigos, agradeço todo o apoio e motivação ao longo do Mestrado em Engenharia e Gestão Industrial, com especial carinho para a Sara, companheira de vida, que me apoia em todos os momentos.

RESUMO

Perante a atual competitividade dos mercados globais, garantir a fiabilidade de todos os ativos de uma unidade de produção assume-se como essencial, no sentido de manter a estabilidade de produção com altos níveis de qualidade.

Com o objetivo de prevenir falhas nos equipamentos, as atividades de manutenção são, habitualmente, realizadas quando é expectável que os componentes se encontrem desgastados. Estes procedimentos interrompem os processos de produção com maior frequência do que o estritamente necessário. Para contornar estes aspetos, a manutenção preditiva revela-se como a estratégia mais comum, assumindo o objetivo de otimizar custos e aumentar a disponibilidade dos equipamentos (Mateus *et al.*, 2021).

Com o presente trabalho, pretende-se desenvolver uma análise de fiabilidade de uma prensa de secagem, com o objetivo de avaliar a eficiência das metodologias de manutenção preditiva aplicadas. A variável em análise consiste na intensidade de corrente elétrica, com base numa amostra de dados recolhidos com uma periodicidade de 1 minuto. O processo de limpeza e tratamento de dados revela-se a tarefa mais desafiante neste processo.

O estimador de *Kaplan-Meier* para a curva de sobrevivência dos tempos de funcionamento da prensa (Kaplan & Meier, 1958) é apresentado ao longo do presente documento. É realizada a comparação de curvas de sobrevivência para diferentes níveis de variáveis explicativas, também consideradas. Por fim, é construído um modelo de regressão de Cox para avaliar a sua influência na função de risco.

Palavras-chave: Manutenção preditiva, análise de sobrevivência, falha.

ABSTRACT

Given the current competitiveness of the global market, it is essential to ensure the reliability of all assets present in a production unit, maintaining manufacturing stability and meeting the required quality.

Maintenance procedures occur when parts are expected to be worn out, preventing failures and halting the production processes for more time than strictly necessary, focusing on preventing future failures. Predictive maintenance is the most common type of approach, aiming to optimize costs and increase equipment availability (Mateus *et al.*, 2021).

With this work, it is intended to conduct a reliability analysis of a pulp drying press, to evaluate the efficiency of the predictive maintenance methodologies applied. The variable monitored is the electric current intensity, considering data gathered with a sampling period of 1 minute. Data cleaning reveals to be most challenging task.

The Kaplan-Meier estimate of the survival function (Kaplan & Meier, 1958) is presented. A comparison of survival curves for different levels of explanatory variable is also presented, as well as a Cox regression model to evaluate their impact in the hazard function.

Keywords: Predictive maintenance, survival analysis, failure.

ÍNDICE

Agradecimentos	i
Resumo	ii
Abstract.....	iii
Índice.....	1
Índice de figuras.....	3
Índice de tabelas	4
Lista de siglas e acrónimos	5
Lista de símbolos	6
1 Introdução	7
1.1 Enquadramento	7
1.2 Objetivos	7
1.3 Manutenção – Considerações	8
1.4 Processo de produção de (pasta de) papel	10
1.5 Estrutura do relatório.....	10
2 Enquadramento Teórico.....	11
2.1 Introdução.....	11
2.2 Conceitos básicos	11
2.2.1 Função densidade de probabilidade e função distribuição de probabilidade	11
2.2.2 Função de sobrevivência	12
2.2.3 Função de risco.....	13
2.3 Censura	16
2.4 Estimador de <i>Kaplan-Meier</i>	18
3 Caso de estudo	24
3.1 Introdução.....	24
3.2 Tratamento de dados	25
3.3 Caracterização estatística	30
3.4 Resultados	32
3.4.1 Curvas de sobrevivência e de risco acumulado	32
3.4.2 Análise de covariáveis	33

3.4.3	Comparação das curvas de sobrevivência	38
3.4.4	Modelos de regressão.....	40
4	Conclusões e propostas futuras	43
	Referências bibliográficas	45

ÍNDICE DE FIGURAS

Figura 2-1 – Variação da taxa de falha ao longo do tempo (Motulsky, 2023).....	13
Figura 2-2 – Diferentes tipos de curvas de risco (adaptado de Ramos (2018))......	16
Figura 2-3 – Representação gráfica de diferentes tipos de censura (Borges, 2014).18	
Figura 2-4 – Exemplo de uma função de sobrevivência (adaptado de DATAtab Team (2023)).....	19
Figura 2-5 – Exemplo de curva de sobrevivência sem censura (dados provenientes de DATAtab Team (2023))......	21
Figura 2-6 – Exemplo de curva de sobrevivência com censura (dados provenientes de DATAtab Team (2023)).....	23
Figura 3-1 – Visualização gráfica dos registos válidos no ano de 2017.....	27
Figura 3-2 – Visualização gráfica dos registos válidos referentes à velocidade de rotação da prensa.....	29
Figura 3-3 – Visualização gráfica dos registos válidos referentes à temperatura.....	30
Figura 3-4 – Visualização gráfica dos registos válidos referentes ao torque.....	30
Figura 3-5 – Histograma dos tempos de sobrevivência.....	31
Figura 3-6 – <i>Boxplot</i> referente aos dados obtidos para os tempos de vida.....	32
Figura 3-7 – Curva de sobrevivência com base nos tempos de vida identificados. 32	
Figura 3-8 – Curva de risco acumulado com base nos tempos de vida identificados.	33
Figura 3-9 – Visualização gráfica dos registos referentes à velocidade no instante anterior a cada paragem.....	34
Figura 3-10 – Visualização gráfica dos registos referentes à temperatura no instante anterior a cada paragem.....	34
Figura 3-11 – Visualização gráfica dos registos referentes ao torque no instante anterior a cada paragem.....	34
Figura 3-12 – Curva de sobrevivência com base na velocidade de rotação.....	35
Figura 3-13 – Curva de risco acumulado com base na velocidade de rotação.	35
Figura 3-14 – Curva de sobrevivência com base na temperatura.....	36
Figura 3-15 – Curva de risco acumulado com base na temperatura.....	37
Figura 3-16 – Curva de sobrevivência com base o torque.....	37
Figura 3-17 – Curva de risco acumulado com base no torque.....	38

ÍNDICE DE TABELAS

Tabela 2.1 – Exemplo de cálculo do estimador <i>Kaplan-Meier</i> : dados sem censura (DATAtab Team, 2023).....	20
Tabela 2.2 – Exemplo de cálculo do estimador <i>Kaplan-Meier</i> : probabilidade de sobrevivência sem censura (adaptado de DATAtab Team (2023)).....	20
Tabela 2.3 – Exemplo de cálculo do estimador <i>Kaplan-Meier</i> : dados com censura (DATAtab Team, 2023).....	21
Tabela 2.4 – Exemplo de cálculo do estimador <i>Kaplan-Meier</i> : registo de falhas/censura (DATAtab Team, 2023).....	22
Tabela 2.5 – Exemplo de cálculo do estimador <i>Kaplan-Meier</i> : probabilidade de sobrevivência com censura (adaptado de DATAtab Team (2023)).....	22
Tabela 3.1 – Descrição dos sensores instalados na prensa em análise.....	24
Tabela 3.2 – Excerto do ficheiro de dados relativo ao sensor 46II015.DACA.PV.24	
Tabela 3.3 – Excerto inicial do <i>dataframe</i> de dados devidamente classificados.	26
Tabela 3.4 – Excerto inicial do <i>dataframe</i> de dados com confiança.	26
Tabela 3.5 – Excerto do <i>dataframe</i> com a informação sobre funcionamento a prensa.	27
Tabela 3.6 – Excerto do <i>dataframe</i> com a informação sobre tempos de vida e de paragem da prensa.	28
Tabela 3.7 – Tempos de paragem superiores a seis dias.....	28
Tabela 3.8 – Excerto inicial do <i>dataframe</i> relativo aos tempos de vida e respetiva classificação de paragem.	29
Tabela 3.9 – Registo de leituras das variáveis em função do intervalo de maior incidência.	35
Tabela 3.10 – Resultados do teste de <i>Log-Rank</i>	39
Tabela 3.11 – Modelo de regressão de <i>Cox</i>	41

LISTA DE SIGLAS E ACRÓNIMOS

IA	Inteligência Artificial
IPQ	Instituto Português da Qualidade
NP-EN	Norma Portuguesa – European Norm

LISTA DE SÍMBOLOS

Alfabeto latino

$\hat{S}_{KM}(t)$	Função de sobrevivência estimada por <i>KM</i>
$\hat{S}(t)$	Função de sobrevivência estimada
b_j	Coefficiente associado à covariável j do modelo de regressão de <i>Cox</i>
C_i	Dados censurados
d_j	Número de falhas no instante j
dt	Intervalo de tempo infinitesimal
E_t	Número de falhas
E_t	Número de eventos esperado
$f(t)$	Função de densidade ou taxa instantânea de falha
$F(t)$	Função de distribuição de probabilidade
$h(t)$	Função de risco
$H(t)$	Função de risco acumulado
H_0	Hipótese nula
$h_0(t)$	Varição da função de risco no tempo de sobrevivência
H_1	Hipótese de teste
<i>IQR</i>	Intervalo entre quartis
<i>KM</i>	Estimador <i>Kaplan-Meier</i>
n	Número de observações
n_j	Observações sem falha no instante j
N_t	Observações em risco
O_t	Número de falhas observadas
P	Probabilidade
Q_1	Primeiro quartil
Q_3	Terceiro quartil
$S(t)$	Função / probabilidade de sobrevivência
T	Tempo decorrido desde o início do estudo
t	Instante aleatório no tempo
t_j	Tempo no instante j
X_j	Covariável j do modelo de regressão de <i>Cox</i>

Alfabeto grego

β_j	Coefficiente associado à covariável j do modelo de regressão de <i>Cox</i>
χ^2	Estatística de teste

1 INTRODUÇÃO

1.1 Enquadramento

Numa altura em que se verificam avanços significativos no desenvolvimento de diversas ferramentas de Inteligência Artificial (IA), dando-lhes grande destaque mediático, torna-se evidente que este tipo de soluções pode representar um forte impulsionador da indústria, impelindo as organizações que tenham condições e decidam canalizar investimento para esta área. Por outro lado, é importante assinalar que, apesar deste paradigma apresentar décadas de desenvolvimento, não deixa de ser relativamente recente no seio da generalidade do meio empresarial e industrial. A forte disseminação de ferramentas baseadas em inteligência artificial de fácil acesso e utilização através da internet, pode transmitir uma sensação de grande simplicidade. No entanto, e como referido, estas soluções representam o resultado de muitos anos de investigação e baseiam-se em sistemas muito complexos como sejam, por exemplo, modelos de redes neuronais (Mateus *et al.*, 2021) ou abordagens gráficas (Xia *et al.*, 2022). Apesar de poder parecer evidente que a implementação deste tipo de modelos em contexto industrial possa representar uma grande vantagem competitiva, proporcionando grandes retornos financeiros e produtivos, é importante que esta transição seja realizada de forma sustentada, o que requer fortes investimentos, não apenas em recursos humanos qualificados, como em equipamentos informáticos capazes de processar este tipo de modelos em tempo útil. Este tipo de investimento poderá, facilmente, ganhar proporções financeiras e temporais muito significativas, tornando-se inviável para muitas organizações.

Perante este cenário, o recurso a ferramentas de implementação menos complexa revela-se uma necessidade. Neste sentido, a aplicação de ferramentas estatísticas que, apesar de também se basearem em cálculos matemáticos exigentes, pode possibilitar uma aplicação mais simples e rápida, representam uma opção viável não apenas para organizações que não tenham possibilidade de investir em modelos mais complexos, mas também como um modo de transição no desenvolvimento desse tipo de sistemas.

Apesar de os resultados das ferramentas de IA se apresentarem como muito promissores, não significa que outras ferramentas não sejam igualmente capazes de apresentar resultados fiáveis e, em muitos casos, perfeitamente adequados aos objetivos das organizações, permitindo extrair informação dos dados recolhidos, sem necessidade de investimentos demasiado avultados.

1.2 Objetivos

Perante a realidade descrita anteriormente, considerou-se pertinente estudar a viabilidade de aplicação de outro tipo de modelos ou ferramentas, que possibilitem a sua implementação com recurso a investimentos reduzidos, mas que permitam

chegar a conclusões relevantes no suporte da tomada de decisões estratégicas e operacionais.

Neste sentido, o trabalho que se apresenta assume-se como um esboço inicial de uma ferramenta de fácil implementação para o estudo do tempo de funcionamento de uma prensa de secagem na indústria da pasta de papel.

1.3 Manutenção – Considerações

Em inúmeras situações, a indústria produtiva caracteriza-se pela necessidade de continuidade de produção, de modo a maximizar a produtividade, evitando períodos de paragem e arranque.

Para este efeito, as boas práticas no meio industrial baseiam-se em elaborados planos de manutenção, que têm como objetivo principal a otimização das linhas de produção. Para alcançar este objetivo, por um lado é mais ou menos evidente que se pretendem reduzir os tempos de paragem das linhas de produção e, por outro lado não se podem negligenciar atividades de manutenção, tais como limpeza de equipamentos, substituição de componentes de desgaste, verificação e nivelamento de sistemas de lubrificação, entre outras.

Existem casos em que muitas destas atividades podem ser realizadas sem ser necessário recorrer à paragem das linhas de produção. Adicionalmente, existem atualmente muitas soluções de monitorização (com recurso a sensores) que permitem, por vezes em tempo real, o acompanhamento do desempenho dos equipamentos. Estas opções permitem reduzir significativamente os tempos de paragem. No entanto, é importante referir que prolongar os tempos de produção poderá aumentar o risco de falha nos equipamentos. Esta situação apresenta-se como o pior cenário possível, uma vez que poderá resultar em situações imprevistas, com consequências potencialmente graves para os equipamentos, que poderão culminar em tempos de paragem ainda mais prolongados.

Existem diversas opções de abordagem para este tipo de situações. Do ponto de vista da manutenção, o decisor poderá optar pela implementação de soluções de manutenção corretiva, preventiva ou preditiva. Como seria de esperar, cada uma destas opções apresenta vantagens e desvantagens.

Segundo a NP-EN 13306 (IPQ, 2007), a manutenção corretiva caracteriza-se pela realização de intervenções que têm o objetivo de repor um equipamento em funcionamento após a deteção de uma falha. Neste contexto, é assumido e consequentemente aceite que irá ocorrer uma falha. Como referido anteriormente, tem a desvantagem de aumentar o risco de danos graves e consequentes tempos de paragem mais longos (Xia *et al.*, 2022).

A manutenção preventiva baseia-se na realização de intervenções a intervalos de tempo pré-definidos ou de acordo com critérios prescritos, com o objetivo de reduzir a probabilidade de avaria do equipamento (NP-EN 13306 (IPQ, 2007)).

Comparativamente à manutenção corretiva, esta apresenta a grande vantagem de permitir acompanhar o estado dos equipamentos ao longo do tempo. Contudo, baseia-se na substituição de componentes que poderiam ainda cumprir a sua função por mais tempo (Xia *et al.*, 2022).

De acordo com a mesma norma (NP-EN 13306 (IPQ, 2007)), a manutenção preditiva integra-se no conceito de manutenção condicionada, que é definida como uma estratégia de manutenção preventiva baseada na vigilância do funcionamento do equipamento. Neste sentido, a norma define manutenção preditiva como uma estratégia de manutenção condicionada que se baseia na realização de intervenções com base na extrapolação de parâmetros monitorizados ao longo do funcionamento do equipamento. Desta forma, uma boa estratégia de manutenção preventiva revela-se uma solução promissora, uma vez que tem o potencial de atingir os benefícios das duas estratégias anteriores. Esta opção combina a monitorização dos equipamentos com a realização de intervenções antes da ocorrência de falha. Por via de uma boa monitorização, a implementação de uma solução de manutenção preditiva permite identificar o momento mais adequado para a realização das intervenções. Neste sentido é possível reduzir a frequência de intervenções bem como a substituição de componentes que se encontram com bons níveis de funcionamento, enquanto evita a ocorrência de falhas (Xia *et al.*, 2022).

No campo da estatística, a análise de sobrevivência representa uma ferramenta que tem como objetivo, entre outros, o de estimar o tempo de funcionamento de determinado equipamento até à ocorrência de uma falha. Este tipo de análise é utilizado em diversas áreas, em especial no ramo da medicina, onde este tipo de informação é crucial para o desenvolvimento e melhor entendimento de tratamentos farmacológicos, por exemplo. Este tipo de análise é amplamente aplicável ao contexto industrial uma vez que permite prever o tempo de vida expectável de um determinado equipamento, com base em dados históricos.

Poderá afirmar-se que a análise de sobrevivência em contexto industrial se enquadra numa filosofia de manutenção preditiva. A estimação do tempo de sobrevivência de um equipamento industrial permite prever o momento de falha. Por sua vez, este tipo de informação tanto pode ser utilizado na elaboração de planos de manutenção, como em contexto de produção, mediante a análise dos dados de monitorização dos equipamentos.

A metodologia de análise de sobrevivência é caracterizada pela existência de dados censurados, razão pela qual não se utilizam outras formas de análise de dados. A existência de dados censurados traduz-se numa restrição na recolha dos dados, que se verifica quando o acontecimento em estudo não ocorre em todos os indivíduos observados. Esta característica impossibilita o tratamento dos dados através de outras metodologias.

1.4 Processo de produção de (pasta de) papel

Como é do conhecimento comum, a principal matéria-prima utilizada na produção de papel é a madeira. Existem diferentes processos de produção de papel que se podem subdividir em dois grupos principais: processos químicos e processos mecânicos. Existem ainda subgrupos destes processos, bem como processos que combinam as duas filosofias (Araújo, 2010). Qualquer destes processos é bastante complexo e envolve o recurso a equipamentos industriais pesados e longas linhas de produção.

Um dos processos mais utilizados é designado por processo Kraft. Este processo inclui-se no grupo dos processos químicos e é constituído por seis fases principais: preparação da madeira (destroçamento), cozimento, lavagem, branqueamento, secagem, enfardamento (Ramalho, 2019).

Em várias fases do processo, após a cozedura, são utilizadas prensas de secagem que têm o objetivo de reduzir a humidade da pasta de papel ao longo do processo. O presente trabalho centra-se no estudo de uma destas prensas, que se insere na fase de branqueamento. Trata-se de uma prensa de funcionamento rotativo com velocidades reduzidas.

1.5 Estrutura do relatório

O presente documento está organizado em quatro capítulos principais. O primeiro capítulo contempla uma breve introdução onde se apresenta a motivação e os objetivos para o presente estudo, bem como algumas considerações sobre a indústria da pasta de papel. No segundo capítulo é apresentado o enquadramento teórico das ferramentas utilizadas no trabalho. De seguida, o terceiro capítulo apresenta um caso de estudo, elaborado com base em dados reais do funcionamento de uma prensa de secagem, inserida numa linha de produção de pasta de papel. Neste capítulo é detalhado o tratamento de dados realizado, bem como os resultados obtidos. Inclui-se, ainda, a apresentação de alguns conceitos teóricos considerados relevantes para a análise. Por fim, o quarto capítulo contempla as conclusões que foi possível retirar do trabalho, bem como algumas propostas para trabalhos futuros.

2 ENQUADRAMENTO TEÓRICO

2.1 Introdução

Como tem sido referido, o presente trabalho pretende avaliar a aplicabilidade de ferramentas estatísticas de fácil utilização em contexto industrial, em concreto na indústria da pasta de papel. Uma solução deste tipo de ferramentas com grande potencial consiste em realizar uma análise de sobrevivência.

A metodologia de análise de sobrevivência consiste na observação e registo do tempo que decorre desde o início de um determinado estudo até à ocorrência de um acontecimento pré-definido. Esta metodologia caracteriza-se pela existência de dados censurados, razão pela qual não se utilizam outras formas de análise de dados. A existência de dados censurados traduz-se numa restrição na recolha dos dados, que se verifica quando o acontecimento em estudo não ocorre em todos os indivíduos observados.

Como o próprio nome poderá sugerir, a metodologia de análise de sobrevivência está muito ligada à área da saúde. Neste campo, a análise de sobrevivência é muito utilizada para estudar a aplicação de tratamentos inovadores, uma vez que permite avaliar a reincidência de uma amostra de indivíduos após a administração do tratamento em estudo. Também no âmbito da sociologia, esta metodologia permite avaliar, por exemplo, o tempo de reincidência de indivíduos com problemas de adição. No meio industrial, a análise de sobrevivência permite estudar o tempo de vida até à substituição de um componente, por exemplo, ou o tempo que decorre até à sua falha, após este ter sido substituído.

2.2 Conceitos básicos

2.2.1 Função densidade de probabilidade e função distribuição de probabilidade

Em análise de sobrevivência, a variável resposta corresponde ao tempo decorrido desde o início do estudo até à ocorrência do acontecimento de interesse (falha). Esta variável, que será designada por T , é aleatória, contínua e não negativa. Como referido, T descreve o tempo de sobrevivência de um determinado indivíduo, ou seja, T representa o tempo decorrido desde o início do estudo até à observação de uma falha. Apesar de os valores de T serem desconhecidos, esta variável pode ser definida através de uma função de densidade representada por $f(t)$ e por uma função de distribuição de probabilidade, representada por $F(t)$ (Motulsky, 2023).

A função de densidade é definida como a probabilidade da variável em estudo (neste caso T), assumir qualquer valor aleatório, t . Por outras palavras, a função de densidade de probabilidade, $f(t)$ representa a taxa instantânea de falha nesse instante sendo dada por (Borges, 2014):

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt)}{dt} \quad (2.1)$$

Por sua vez, a função distribuição é definida como a probabilidade da variável em estudo (T) ser inferior ou igual a um valor aleatório, t (Singh, 2023). Matematicamente, a função distribuição pode ser definida por:

$$F(t) = P(T \leq t), 0 \leq t < \infty \quad (2.2)$$

Estas funções apresentam algumas particularidades que ajudam na sua interpretação. Em primeiro lugar, o valor de $f(t)$ é maior ou igual a zero para todos os valores de t . Por definição, a área sob a curva definida por $f(t)$ é igual à unidade e a função distribuição pode ser obtida pelo integral da função densidade (Motulsky, 2023). Estas propriedades podem ser definidas matematicamente através das seguintes formulações:

$$\int_0^{\infty} f(t) = 1 \quad (2.3)$$

$$F(t) = \int_0^t f(u) du \quad (2.4)$$

2.2.2 Função de sobrevivência

A função de sobrevivência da variável em estudo (T) consiste na probabilidade de um determinado evento não ter ocorrido até um instante aleatório, t . Com base nas propriedades descritas anteriormente é possível concluir que a probabilidade de um determinado evento não ter ocorrido até um instante aleatório, t , consiste na diferença entre a unidade e o valor da função distribuição, $F(t)$:

$$P(T > t) = 1 - F(t) \quad (2.5)$$

Substituindo as expressões (2.3) e (2.4) na expressão (2.5), obtém-se a definição matemática da função de sobrevivência (Motulsky, 2023):

$$P(T > t) = 1 - F(T) = \int_0^{\infty} f(u) du - \int_0^t f(u) du = \int_t^{\infty} f(u) du \quad (2.6)$$

$$S(t) = \int_t^{\infty} f(u) du \quad (2.7)$$

De forma simplificada, a função de sobrevivência pode ser reescrita pela seguinte expressão:

$$S(t) = P(T > t), \forall t \geq 0 \quad (2.8)$$

Esta função é monótona, não crescente e contínua, e caracteriza-se pelas seguintes propriedades:

$$S(0) = 1 \quad (2.9)$$

$$S(+\infty) = \lim_{t \rightarrow \infty} S(t) = 0 \quad (2.10)$$

2.2.3 Função de risco

O tempo de vida dependerá sempre de alguns fatores externos. A ocorrência de uma falha, por exemplo, será o suficiente para que este termine. Como tal, é necessário considerar as funções de risco, $h(t)$, e de risco acumulado, $H(t)$. A função de risco (também designada por taxa de risco) pode ser vista como a frequência de ocorrência de falhas por unidade de tempo, considerando que não ocorreu falha até esse instante. Desta forma, uma taxa de risco elevada significa maior número de falhas num determinado instante. Menor taxa de falha significa menos falhas nesse instante. Como ilustrado na Figura 2-1, a taxa de falha poderá variar ao longo do tempo (Motulsky, 2023).

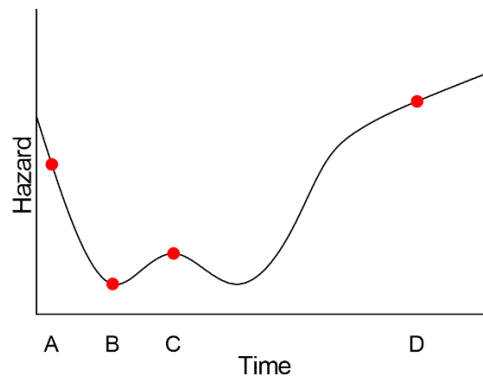


Figura 2-1 – Variação da taxa de falha ao longo do tempo (Motulsky, 2023).

No caso de se tratar de uma variável discreta, a taxa de risco pode ser definida como a probabilidade de a falha ocorrer num instante específico, dado que não ocorreu anteriormente (Motulsky, 2023):

$$h(t) = P(T = t | T \geq t) \quad (2.11)$$

No caso de variáveis contínuas, a formulação torna-se mais complexa. Considere-se um intervalo de tempo definido por $[t1, t2[$. A probabilidade de ocorrer uma falha neste intervalo é definida em termos da função $S(t)$, na forma $S(t1) - S(t2)$. O quociente entre a probabilidade de ocorrência da falha, no intervalo de tempo $[t1, t2[$, e a amplitude do intervalo, representa a taxa de falha.

Considerando um intervalo de tempo de vida tão pequeno quanto se pretenda, obtém-se a taxa instantânea de falha para o instante t , que representa a função de risco $h(t)$. Neste caso, t representa o valor final do intervalo do tempo de vida. Desta

forma, a função de risco é definida pela equação (2.12) e caracteriza-se pelas condições (2.13) e (2.14) (Borges, 2014):

$$h(t) = \lim_{dt \rightarrow 0^+} \left(\frac{P(t \leq T < t + dt | T \geq t)}{dt} \right) \quad (2.12)$$

$$h(t) \geq 0 \quad (2.13)$$

$$\int_0^{\infty} h(u) du = \infty \quad (2.14)$$

A função de risco, $h(t)$, poderá fornecer informação complementar à função de sobrevivência. Se, por um lado, diferentes funções de sobrevivência podem ter desenvolvimentos gráficos semelhantes, os gráficos da função de risco podem diferir significativamente.

Considerando as definições anteriores, é possível reescrever a função de risco de forma simplificada (Borges, 2014):

$$h(t) = \frac{f(t)}{S(t)} \quad (2.15)$$

Nesta formulação, a função de risco pode ser vista como a densidade de falhas num determinado instante t , dividido pela probabilidade de não ter ocorrido falha até esse instante.

Outro conceito importante na implementação de análise de sobrevivência é a função de risco acumulado. Este parâmetro pode ser visto como o risco de ocorrência de falha que se acumulou ao longo do tempo. Ao contrário da taxa de risco instantânea, a função de risco acumulado apenas aumenta ao longo do tempo. Por outras palavras, quanto mais tempo decorrer sem se verificar uma falha, mais provável se torna a sua ocorrência. Do ponto de vista matemático, isto explica-se pelo facto da função de risco instantânea ser não negativa (Motulsky, 2023).

A função de risco acumulado pode ser obtida através do integral da taxa de risco instantânea, ou seja:

$$H(t) = \int_0^t h(u) du, t \geq 0 \quad (2.16)$$

Por outras palavras, a função de risco acumulado mede o risco de ocorrência de falha até um determinado instante (Borges, 2014).

Para compreender a relação entre a função de risco acumulado e a função de sobrevivência, observe-se a definição da função de sobrevivência apresentada anteriormente (2.6). A função de sobrevivência pode ser escrita na seguinte formulação:

$$S(t) = 1 - \int_0^t f(u)du \quad (2.17)$$

Derivando ambos os lados da equação (2.17) obtém-se:

$$\frac{d}{dt}S(t) = -f(t) \quad (2.18)$$

Aplicando um logaritmo ao termo esquerdo da equação anterior, é possível tirar partido das propriedades matemáticas dos logaritmos. Desta forma, sabendo que a derivada de um logaritmo corresponde à derivada do seu argumento sobre o próprio argumento, obtém-se (Motulsky, 2023):

$$\frac{d}{dt}\ln(S(t)) = \frac{d}{dt}S(t) * \frac{1}{S(t)} \quad (2.19)$$

Substituindo (2.18) na expressão (2.19) obtém-se o inverso da função de risco apresentada anteriormente:

$$\frac{d}{dt}\ln(S(t)) = \frac{-f(t)}{S(t)} \quad (2.20)$$

$$\frac{d}{dt}\ln(S(t)) = -h(t) \quad (2.21)$$

Ao aplicar a derivada do exponencial da expressão anterior, obtém-se a função de sobrevivência, definida através da função de risco:

$$S(t) = \exp\left(-\int_0^t h(x)dx\right) \quad (2.22)$$

Esta equação permite estabelecer a relação entre a função de sobrevivência e a função de risco acumulado. Esta relação pode ser rescrita simplificadamente através das seguintes formulações:

$$S(t) = \exp(-H(t)) \quad (2.23)$$

$$H(t) = -\ln(S(t)) \quad (2.24)$$

A natureza decrescente da função de sobrevivência poderá dificultar a perceção de como evolui o risco de falha ao longo do tempo. Por outro lado, a observação do gráfico definido pela função de risco evidencia a evolução do risco de falha ao longo do tempo (Borges, 2014). Podem destacar-se cinco tipos de comportamentos mais comuns, como ilustrado na Figura 2-2 (Ramos, 2018):

- Monótona decrescente (*decreasing*): representa a situação menos frequente em que o risco de falha diminui ao longo do tempo. Estão habitualmente

associadas a situações em que eventuais defeitos de fabrico representam a principal causa de falha;

- Monótona crescente (*increasing*): em oposição à situação anterior, é um comportamento muito comum que traduz situações em que se verifica um envelhecimento gradual. Um dos exemplos mais evidentes consiste em estudos médicos de casos terminais ou doenças incuráveis, onde a longevidade se associa a um maior risco de morte;
- Unimodal (*bump-shaped*): reflete situações em que o risco de falha começa por aumentar até um valor máximo, a partir do qual começa a decrescer. Está, por exemplo, associado a estudos de sobrevivência após cirurgias bem-sucedidas;
- Constante: situação particular em que o risco de falha não varia durante o estudo. Habitualmente verifica-se em estudos onde o período de observação é muito reduzido ou em que ocorre uma situação imprevista que determina o fim abrupto do estudo;
- Em forma de “U” ou banheira (*Bathtub / U-shaped*): caracteriza-se por comportamentos em que o risco inicial é elevado, seguidos de um longo período de risco constante com o aumento no final de vida. Na maioria dos casos implica um período de observação desde o início de vida, altura em que há mais risco de falha devido a defeitos de fabrico. Depois de corrigidos os defeitos iniciais, o período de vida útil apresenta um risco de falha constante que posteriormente aumenta devido ao desgaste de utilização. Outro exemplo consiste na vida humana, que apresenta maior risco de morte em recém-nascidos, bem como na fase final da vida.

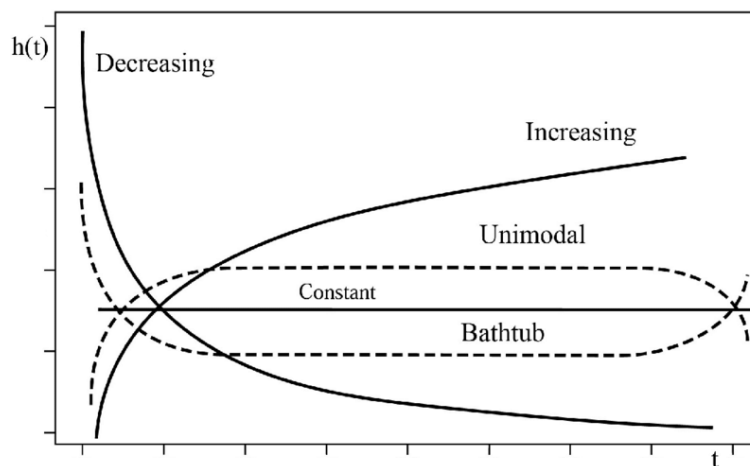


Figura 2-2 – Diferentes tipos de curvas de risco (adaptado de Ramos (2018)).

2.3 Censura

Quando se pretende estudar dados relacionados com sobrevivência (tanto no campo da medicina como na área industrial) é comum ocorrerem situações em que alguma

informação dos participantes do estudo seja incompleta. Nestes casos estamos perante aquilo que se designa como dados censurados. A abordagem estatística de análise de sobrevivência permite ter em consideração este tipo de situações, o que representa uma grande vantagem em comparação com outras abordagens, que não dispõem de ferramentas para esse efeito.

A censura pode ocorrer devido a, essencialmente, três causas (Borges, 2014):

- A não observação de um acontecimento de interesse durante o estudo – Por exemplo situações de falha ou óbito devido a causa não relacionada com o estudo ou o estudo atingir o final do período de observação sem se ter verificado o acontecimento de interesse;
- Perda de informação de continuidade de um indivíduo ao longo do estudo – Mudança de residência de um indivíduo (no caso de estudos médicos/sociológicos) ou substituição do equipamento industrial por motivos não relacionados ao estudo;
- A remoção de um dos indivíduos do estudo – Mudança de procedimento industrial ou efeitos adversos, no caso de indivíduos.

Uma variável aleatória é caracterizada como censurada quando não se consegue definir o seu valor exato, mas é possível obter um limite inferior (censura à direita), ou limite superior (censura à esquerda). No caso de se verificarem ambas as situações, tratar-se-á de censura intervalar. Estas situações de censura podem ser caracterizadas em cinco tipos diferentes (Borges, 2014):

- Censura à direita: Verifica-se quando o tempo de vida de um determinado indivíduo é superior a um determinado valor. Este tipo de censura pode ainda ser subdividido em três subtipos:
 - Censura de tipo I: Nos casos em que o instante final para o estudo é previamente fixado e alguns elementos ainda não registaram falha nesse momento;
 - Censura de tipo II: Se, alternativamente, é fixado o número de falhas do acontecimento de interesse para ditar o final do estudo. Sendo que o número de falhas pré-definido será inferior ao número de elementos no estudo;
 - Censura aleatória: Ocorre quando um elemento é retirado no decorrer do estudo sem ter sido registada a falha;
- Censura à esquerda: acontece quando o evento de interesse tenha ocorrido num determinado instante de tempo anterior à observação, ou seja, o tempo de vida é inferior ao tempo em estudo;
- Censura intervalar: Não se consegue identificar o momento em que ocorreu o evento de interesse, apenas se sabe o intervalo de tempo ao qual pertence;

- Censura independente: ocorre quando a razão para a censura é independente da razão da falha;
- Censura não informativa: ocorre quando se garante que um indivíduo sobre o qual ocorre a censura é representativo de todos os restantes sobreviventes e cujas mesmas características ou covariáveis são as mesmas.

A representação gráfica pode ajudar a interpretação destas situações. A Figura 2-3 representa graficamente um exemplo de um estudo de análise de sobrevivência com a duração de 12 anos. É possível observar alguns exemplos de situações que podem ocorrer neste tipo de estudos. No primeiro caso, verifica-se o acontecimento de interesse ao ano 6, pelo que este registo não se encontra censurado. O indivíduo B não regista falha durante todo o período do estudo, pelo que se trata de censura à direita do tipo 1. Os indivíduos C e E apenas entram no estudo no ano 2 mas ambos são removido antes do final do estudo. Estes casos constituem censura aleatória. Por fim o caso D, que também entra no estudo depois do seu início, não verifica falha até ao final do estudo (Borges, 2014).

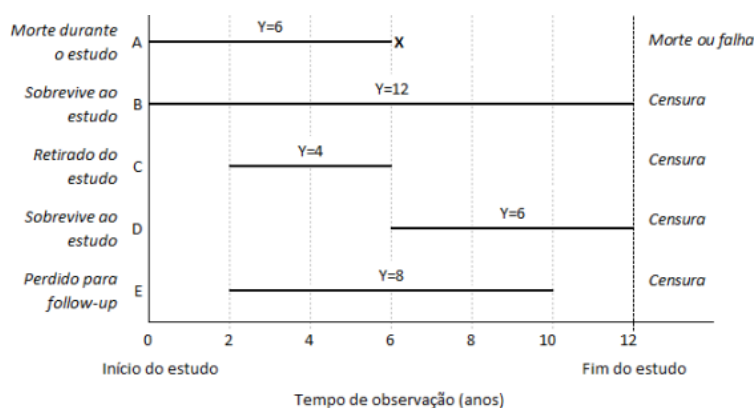


Figura 2-3 – Representação gráfica de diferentes tipos de censura (Borges, 2014).

2.4 Estimador de Kaplan-Meier

Como tem sido referido, o principal objetivo da análise de sobrevivência consiste na obtenção da função de sobrevivência. No caso de não se verificar a existência de dados censurados, a função de sobrevivência $\hat{S}(t)$ seria definida pelo quociente entre o número de observações sem falha até ao instante t e o número total de observações. Desta forma, a função de sobrevivência pode ser representada graficamente sob a forma de escada, como ilustrado na Figura 2-4.

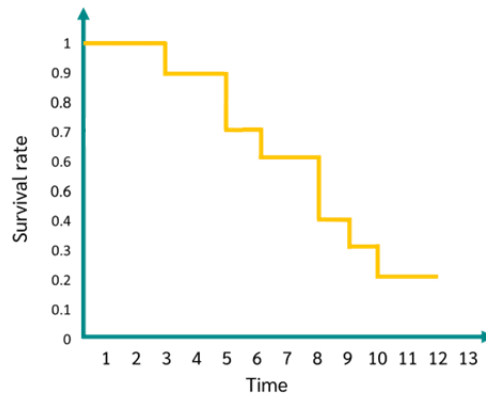


Figura 2-4 – Exemplo de uma função de sobrevivência (adaptado de DATAtab Team (2023))

As descontinuidades representadas no gráfico correspondem aos diferentes tempos de sobrevivência observados. A dimensão de cada descontinuidade tem a dimensão $1/n$ multiplicado pelo número de vezes que se observou cada tempo de vida. A variável n corresponde ao número de observações.

A existência de censura torna o processo mais complexo, impedindo o seu tratamento através de abordagens descritivas convencionais. Desta forma, é necessário recorrer a estimadores não paramétricos, tais como o estimador *Kaplan-Meier*. Este estimador também pode ser designado por estimador limite-produto e é definido como a probabilidade de sobrevivência por um determinado período de tempo. Nesta formulação o tempo é subdividido em pequenos intervalos. A aplicação desta técnica tem por base alguns pressupostos (Kishore *et al.*, 2010):

1. É assumido que os indivíduos que apresentam dados censurados têm a mesma perspetiva de sobrevivência que os indivíduos que se mantêm no estudo;
2. De forma idêntica, é assumido que a probabilidade de sobrevivência é independente do momento de entrada no estudo;
3. Considera-se que o evento de interesse (falha/morte) acontece no momento em que é detetado. Este pressuposto pode envolver alguma falta de exatidão em situações de deteção de falha em controlos periódicos. Nestes casos apenas se sabe que a falha ocorreu entre dois momentos de controlo. Estes casos são mais frequentes em estudos médicos do que no âmbito industrial, uma vez que uma falha (grave) numa linha de produção é imediatamente detetada.

Como referido anteriormente, o cálculo do estimador de *Kaplan-Meier* tem por base a probabilidade de ocorrência de falha num determinado instante. Este valor é obtido em função do instante anterior.

No sentido de exemplificar este cálculo, considere-se um exemplo genérico baseado nos dados da Tabela 2.1 que representam um estudo com 10 observações de tempos de vida (DATAtab Team, 2023). A primeira coluna corresponde ao identificador da

observação; a segunda coluna indica o período que decorreu até à ocorrência de falha, em anos. Numa primeira fase, considere-se que não existem dados censurados.

Tabela 2.1 – Exemplo de cálculo do estimador *Kaplan-Meier*: dados sem censura (DATAtab Team, 2023).

Observações (id)	Tempo decorrido em anos – t
1	3
2	4
3	4
4	6
5	7
6	7
7	8
8	10
9	11
10	13

O primeiro passo consiste em registar o número de falhas em cada intervalo de tempo, bem como o número de observações em risco no início de cada intervalo de tempo. Estes dados são registados na segunda e terceira colunas da Tabela 2.2. Neste caso, a probabilidade de sobrevivência é determinada pelo quociente entre o número de observações em risco (note-se que é necessário subtrair as falhas em cada período) e o número total de observações (DATAtab Team, 2023).

Tabela 2.2 – Exemplo de cálculo do estimador *Kaplan-Meier*: probabilidade de sobrevivência sem censura (adaptado de DATAtab Team (2023)).

Tempo decorrido em anos – t	Número de Falhas – E_t	Obs. em risco N_t	Probabilidade de sobrevivência – $S(t)$	
0	0	10	10/10	1
3	1	10	$(10-1) / 10 = 9 / 10$	0.9
4	2	9	$(9-2) / 10 = 7 / 10$	0.7
6	1	7	$(7-1) / 10 = 6 / 10$	0.6
7	2	6	$(6-2) / 10 = 4 / 10$	0.4
8	1	4	$(4-1) / 10 = 3 / 10$	0.3
10	1	3	$(3-1) / 10 = 2 / 10$	0.2
11	1	2	$(2-1) / 10 = 1 / 10$	0.1
13	1	1	$(1-1) / 10 = 0 / 10$	0

A Figura 2-5 ilustra o gráfico da curva de sobrevivência, obtido através do estimador *Kaplan-Meier*.

Manutenção Preditiva de uma Prensa de Secagem

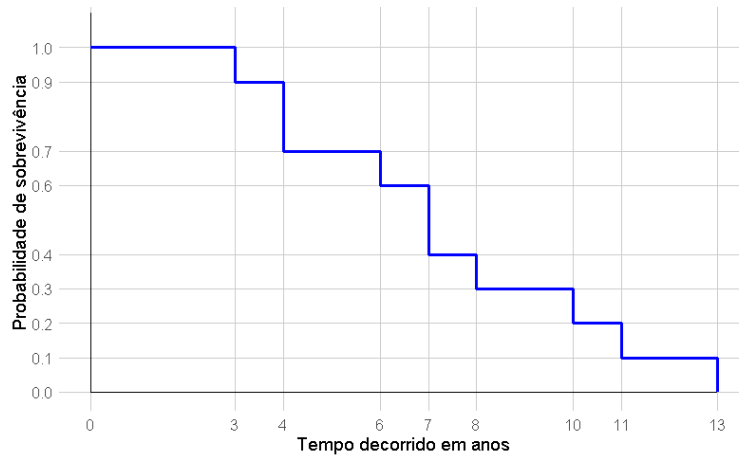


Figura 2-5 – Exemplo de curva de sobrevivência sem censura (dados provenientes de DATAtab Team (2023)).

Não existindo dados censurados, a formulação matemática para estimar a função de sobrevivência pode ser definida por (Borges, 2014):

$$\hat{S}(t) = \frac{\text{Número de tempos de vida} > t}{\text{Número de observações}}, t \geq 0 \quad (2.25)$$

No caso de se verificar a existência de dados censurados será necessário a execução de um passo intermédio para ter este aspeto em consideração.

A Tabela 2.3 repete os dados considerados no exemplo anterior, mas incluindo 3 dados adicionais que foram censurados. Estes dados são assinalados na segunda coluna com o valor 0.

Tabela 2.3 – Exemplo de cálculo do estimador *Kaplan-Meier*: dados com censura (DATAtab Team, 2023).

Tempo decorrido em anos – t	Existência de falha
3	1
4	1
4	1
4	0
6	1
7	1
7	1
8	1
9	0
10	1
11	1
13	1
15	0

De forma idêntica ao exemplo anterior, são registados as falhas e o número de observações em risco no início de cada período. Neste caso, regista-se também o

número de dados censurados em cada intervalo (Tabela 2.4). No instante 0 ainda não ocorreu nenhuma falha nem existem dados censurados. Por outras palavras, todos os casos estão em risco, pelo que a probabilidade de sobrevivência é 1. Para os restantes casos, a probabilidade de sobrevivência é determinada pelo quociente entre o número de observações em risco no final do período e o número de observações em risco no início do período. No entanto, uma vez que existem dados censurados, este cálculo é iterativo, pelo que o quociente anterior é multiplicado pela probabilidade do período antecedente (Tabela 2.5). Note-se que no caso de períodos em que se verifica a ocorrência de censura, mas não se observaram falhas, os dados censurados são contabilizados no período anterior (DATAtab Team, 2023).

Desta forma, a probabilidade de sobrevivência estimada em cada instante pode ser formulada por (Motulsky, 2023):

$$\hat{S}_t = \hat{S}_{t-1} \frac{N_t - E_t}{N_t} \quad (2.26)$$

Tabela 2.4 – Exemplo de cálculo do estimador *Kaplan-Meier*: registo de falhas/censura (DATAtab Team, 2023).

Tempo decorrido em anos – t	Número de Falhas – E_t	Dados censurados – C_t	Obs. em risco – N_t
0	0	0	13
3	1	0	13
4	2	1	12
6	1	0	9
7	2	0	8
8	1	1	6
10	1	0	4
11	1	0	3
13	1	1	2

Tabela 2.5 – Exemplo de cálculo do estimador *Kaplan-Meier*: probabilidade de sobrevivência com censura (adaptado de DATAtab Team (2023)).

Tempo de sobrevivência em anos – t	Probabilidade de sobrevivência – $S(t)$	
0	13/13	1.000
3	$(13-1) / 13 * 1 = 12 / 13 * 1$	0.923
4	$(12-2) / 12 * 0.923 = 10 / 12 * 0.923$	0.769
6	$(9-1) / 9 * 0.769 = 8 / 9 * 0.769$	0.684
7	$(8-2) / 8 * 0.684 = 6 / 8 * 0.684$	0.513
8	$(6-1) / 6 * 0.513 = 5 / 6 * 0.513$	0.428
10	$(4-1) / 4 * 0.428 = 3 / 4 * 0.428$	0.321
11	$(3-1) / 3 * 0.321 = 2 / 3 * 0.321$	0.214
13	$(2-1) / 2 * 0.214 = 1 / 2 * 0.214$	0.107

A Figura 2-6 apresenta a curva de sobrevivência construída com base nestes dados, tendo em conta o efeito da censura.

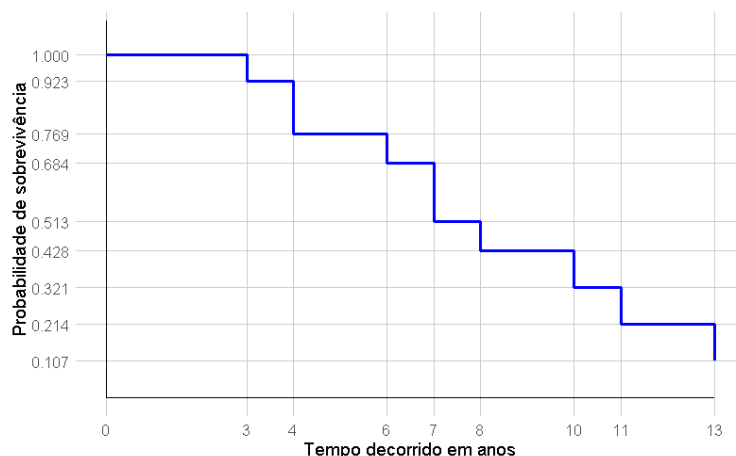


Figura 2-6 – Exemplo de curva de sobrevivência com censura (dados provenientes de DATAtab Team (2023))

Genericamente, considerando t_j como o tempo no instante j em que ocorreu pelo menos uma falha; d_j como o número de falhas observadas no instante t_j e n_j , o número de observações para as quais não se verificou falha, o estimador de *Kaplan-Meier* pode ser definido matematicamente por:

$$\hat{S}_{KM}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) \quad (2.27)$$

3 CASO DE ESTUDO

3.1 Introdução

No sentido de comprovar as potencialidades da aplicação de análise de sobrevivência em contexto real, obtiveram-se dados de uma prensa rotativa, presente numa linha de produção de pasta de papel.

A prensa encontra-se equipada com um sistema de monitorização, composto por diversos sensores nos pontos considerados mais relevantes. A Tabela 3.1 apresenta uma descrição dos sensores instalados na prensa em análise.

Tabela 3.1 – Descrição dos sensores instalados na prensa em análise.

Referência do sensor	Descrição do sensor
46TI084.DACA.PV	Temperatura unidade hidráulica. Prensa nº 1 – 462.016
46LI085.DACA.PV	Nível unidade hidráulica. Prensa nº 1 – 462.016
46SI089.DACA.PV	Velocidade rotação. Prensa nº1 – 462.015
46PC088.DACA.PV	Torque. Prensa nº1 – 462.015
46II015.DACA.PV	Intensidade corrente acionamento sem-fim de extração. Prensa nº1 – 462.015
46PI090.DACA.PV	Pressão VAT cuba. Prensa nº1 – 462.015
46PI094.DACA.PV	Pressão VAT cuba. Prensa nº1 – 462.015
46FC109.DACA.PV	Alimentação pasta. Prensa nº1 – 462.015
46FC110.DACA.PV	Alimentação pasta. Prensa nº1 – 462.015
46FC111.DACA.PV	Alimentação pasta. Prensa nº1 – 462.015
46FC112.DACA.PV	Alimentação pasta. Prensa nº1 – 462.015

Os dados recolhidos são armazenados em diversos ficheiros *.csv* (*comma-separated values*). Cada ficheiro contém informação de um determinado sensor num determinado ano. Os dados disponíveis abrangem o período compreendido entre janeiro de 2017 e outubro de 2020.

A primeira prioridade para a aplicação de uma análise de sobrevivência consiste em identificar os períodos em que o sistema se encontra em funcionamento, bem como os períodos em que o sistema se encontra fora de funcionamento. Para este efeito recorreu-se aos dados relativos ao sensor 46II015.DACA.PV que se referem à intensidade de corrente elétrica. A Tabela 3.2 apresenta um excerto dos dados referentes a este sensor.

Tabela 3.2 – Excerto do ficheiro de dados relativo ao sensor 46II015.DACA.PV.

Times tamp	46II015.DAC A.PV - SnapShot	46II015.DACA.PV - SnapShot - Confidence	Times tamp	46II015.DAC A.PV - SnapShot	46II015.DACA.PV - SnapShot - Confidence
01/01/ 2019	27.59167	100	02/01/ 2019	27.64128	100

Times tamp	46II015.DAC A.PV - SnapShot	46II015.DACA.PV - SnapShot - Confidence	Times tamp	46II015.DAC A.PV - SnapShot	46II015.DACA.PV - SnapShot - Confidence
01/01/ 2019 00:01	27.86099	100	02/01/ 2019 00:01	27.15933	100
01/01/ 2019 00:02	27.6909	100	02/01/ 2019 00:02	27.02467	100
01/01/ 2019 00:03	27.27273	100	02/01/ 2019 00:03	27.20186	100
01/01/ 2019 00:04	27.59167	100	02/01/ 2019 00:04	26.91835	100

A observação da Tabela 3.2 permite concluir que os dados recolhidos representam o momento de cada registo (colunas designadas por *timesamp*), a leitura da intensidade de corrente (colunas designadas pelo código do sensor seguido de *SnapShot*) e colunas referentes ao nível de confiança de cada registo (colunas designadas pelo código do sensor seguido de *SnapShot – Confidence*). Note-se que o excerto apresentado apenas contempla as primeiras linhas bem como as primeiras colunas do ficheiro de registos. Uma análise dos dados de forma mais detalhada permite perceber que o registo é efetuado ao minuto e que cada três colunas correspondem a um mês de registos.

3.2 Tratamento de dados

Com o objetivo de elaborar uma análise de sobrevivência foi necessário realizar diversos processos de tratamento de dados, no sentido de obter uma tabela com os tempos de sobrevivência associados ao tipo de paragem consequente. Para este efeito recorreu-se ao software *RStudio*. Este *software* consiste num ambiente de desenvolvimento informático de utilização livre, baseado na linguagem de programação R. Por sua vez, a linguagem de programação R foi desenvolvida para fins de computação estatística.

Com recurso a estas ferramentas, numa primeira fase foi necessário combinar os dados dos ficheiros correspondentes aos vários anos de registos do sensor relativo à intensidade de corrente. De seguida, procedeu-se a pequenas correções dos dados, tais como a eliminação de valores duplicados (o último valor de cada mês repete-se no início do mês seguinte), ou a correção dos valores iniciais da coluna correspondente ao momento do registo (informação de hora e minuto em falta).

De seguida foi criado um *dataframe*, do qual se apresenta um excerto na Tabela 3.3, com a informação devidamente classificada (dados temporais e numéricos) de modo a permitir a realização de operações matemáticas, bem como a posterior aplicação de funções de análise de sobrevivência.

Tabela 3.3 – Excerto inicial do *dataframe* de dados devidamente classificados.

timestamp	Value	conf
01/01/2017 00:00	146797.8	0
01/01/2017 00:01	146797.2	0
01/01/2017 00:02	146796.8	0
01/01/2017 00:03	146796.3	0
01/01/2017 00:04	146795.8	0
01/01/2017 00:05	146795.3	0

A observação inicial dos dados permitiu, também, identificar diversos registos com valores de intensidade de corrente muito dispare da maioria dos restantes. Na sua generalidade, estes valores correspondem a leituras cuja confiança apresenta valores muito reduzidos. Neste sentido, o passo seguinte passou por remover os dados cuja confiança se considera insuficiente. Para este efeito definiu-se um limite mínimo de confiança de 90, ou seja, todos os registos cuja confiança é inferior a 90 foram removidos da análise. Ao realizar esta operação, a informação contida na coluna referente à confiança torna-se desnecessária, pelo que esta coluna é eliminada. Desta forma, obtém-se um *dataframe* com apenas duas colunas, uma coluna correspondente ao registo temporal e a outra correspondente ao valor de intensidade de corrente (Tabela 3.4).

Tabela 3.4 – Excerto inicial do *dataframe* de dados com confiança.

timestamp	Value
25/07/2017 11:40	24.87597
25/07/2017 11:41	24.88188
25/07/2017 11:42	24.77843
25/07/2017 11:43	24.97874
25/07/2017 11:44	24.75569
25/07/2017 11:45	24.8996

Como referido, na Tabela 3.4 é possível observar os primeiros registos são relativos ao mês de julho de 2017, visto serem os primeiros dados com confiança acima do limite mínimo considerado. Da mesma forma, os valores de intensidade de corrente são muito inferiores aos valores dos primeiros dados desse ano, patentes na Tabela 3.3, que apresentam confiança nula.

Para facilitar a interpretação destes dados, a Figura 3-1 apresenta uma visualização gráfica dos registos considerados válidos, ou seja, dos registos com confiança acima do limite mínimo, no ano 2017. É possível observar que a maioria dos registos se encontram entre o valor de 20 e 30. Verifica-se a existência de alguns picos de corrente, e diversos momentos de corrente nula.

Manutenção Preditiva de uma Prensa de Secagem

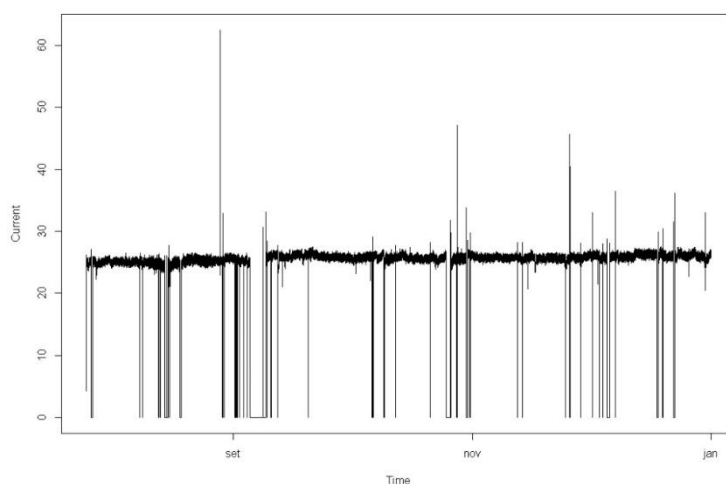


Figura 3-1 – Visualização gráfica dos registos válidos no ano de 2017.

Para identificar os períodos em que a prensa se encontra em funcionamento definiu-se, igualmente, um valor limite de intensidade de corrente, abaixo do qual se considera que o equipamento se encontra fora de funcionamento. O valor considerado foi de 5 e esta informação foi armazenada numa coluna identificada como *onoff*.

De seguida, acrescentou-se uma variável (designada por *status*) com os registos referentes ao arranque e paragem da prensa. Estes registos foram identificados através da comparação com o registo anterior, ou seja, o momento em que o valor armazenado na variável *onoff* se altera de *on* para *off* corresponde a um registo de paragem. De igual forma, quando a variável *onoff* se altera de *off* para *on*, corresponde a um momento de arranque. Na Tabela 3.5 é possível observar um momento em que a intensidade de corrente é inferior ao limite mínimo considerado.

Tabela 3.5 – Excerto do *dataframe* com a informação sobre funcionamento a prensa.

timestamp	value	onoff	status
25/07/2017 11:59	24.78974	on	on
25/07/2017 12:00	24.87715	on	on
25/07/2017 12:01	4.22888	off	stop
25/07/2017 12:02	20.44434	on	start
25/07/2017 12:03	21.89095	on	on

Nesta fase, mantiveram-se apenas os registos relativos aos momentos de paragem e arranque do equipamento. Os restantes registos foram descartados, uma vez que não contêm informação relevante para a análise pretendida.

A diferença entre os momentos de arranque e paragem permitem construir um *dataframe* com os tempos de vida e de paragem do equipamento (Tabela 3.6). A variável *statusNum* contém a mesma informação da variável *status* mas em formato numérico, para facilitar a realização de operações matemáticas. A variável *statusTime* corresponde ao tempo de vida ou de paragem da prensa (em minutos). Por fim, a variável *LD* indica tratar-se de um tempo de vida ou tempo de paragem.

Tabela 3.6 – Excerto do *dataframe* com a informação sobre tempos de vida e de paragem da prensa.

timestamp	value	onoff	status	statusNum	statusTime	LD
25/07/2017 11:40	24.87597	on	on	NA	NA	<NA>
25/07/2017 12:01	4.22888	off	stop	1	21	live
25/07/2017 12:02	20.44434	on	start	0	1	dead
26/07/2017 15:47	3.960736	off	stop	1	1665	live
26/07/2017 17:51	17.20965	on	start	0	124	dead
26/07/2017 18:04	0	off	stop	1	13	live
26/07/2017 23:51	17.01474	on	start	0	347	dead
08/08/2017 01:39	4.160368	off	stop	1	17388	live
08/08/2017 02:14	18.42043	on	start	0	35	dead
08/08/2017 17:11	0.028351	off	stop	1	897	live

No excerto dos dados apresentado na Tabela 3.6 é possível observar um registo de um tempo de paragem de um minuto. Podem existir vários motivos para uma baixa de intensidade de corrente instantânea. Esta situação poderá não consistir numa paragem efetiva dada a sua curta duração. Neste sentido, definiu-se, também, um tempo mínimo de cinco minutos para que uma paragem seja considerada na análise. Os tempos de paragem que não cumprem este critério foram ignorados, considerando-se que o equipamento continua em funcionamento contínuo. Para este efeito, os tempos de paragem inferiores ao limite estipulado foram adicionados aos tempos de vida adjacentes.

A informação disponível relativamente a atividades de manutenção programada do equipamento em análise indica que estas têm uma duração aproximada superior a seis dias e que são pouco frequentes (inferior a uma paragem programada por ano). No período em análise observam-se duas paragens que cumprem este critério (Tabela 3.7), em junho de 2019 e em outubro de 2020.

Tabela 3.7 – Tempos de paragem superiores a seis dias.

timestamp	statusTime	statusNum
01/06/2019 15:28	17935	0
21/10/2020 19:10	14356	0

Com base nesta informação, os tempos de vida que antecedem estas paragens constituem dados censurados, uma vez que o equipamento parou sem que se tenha verificado uma falha. As restantes paragens constituem algum tipo de falha do equipamento. No sentido de realizar uma análise de sobrevivência construiu-se um *dataframe* com os tempos de vida e a respetiva classificação de falha ou censura, respetivamente 1 ou 0 (Tabela 3.8). Para facilitar a interpretação de resultados, a variável *statusTime* é apresentada em dias.

Manutenção Preditiva de uma Prensa de Secagem

Tabela 3.8 – Excerto inicial do *dataframe* relativo aos tempos de vida e respetiva classificação de paragem.

timestamp	statusTime	type
25/07/2017 12:01	1.171528	1
26/07/2017 18:04	0.009028	1
08/08/2017 01:39	12.075	1
08/08/2017 17:11	0.622917	1
12/08/2017 20:02	4.666667	1
14/08/2017 09:33	0.926389	1

O tratamento de dados realizado até este ponto seria suficiente para realizar uma análise de sobrevivência, contudo, a consideração de outras variáveis poderá permitir a obtenção de mais informação sobre o funcionamento da prensa. Neste sentido, cruzaram-se os dados referentes aos tempos de vida com os dados relativos aos sensores relativos à velocidade da prensa, temperatura e torque.

As Figuras 3-2 a 3-4 apresentam os gráficos com todos os registos dos sensores referidos, cuja confiança é superior ao limite mínimo estabelecido.

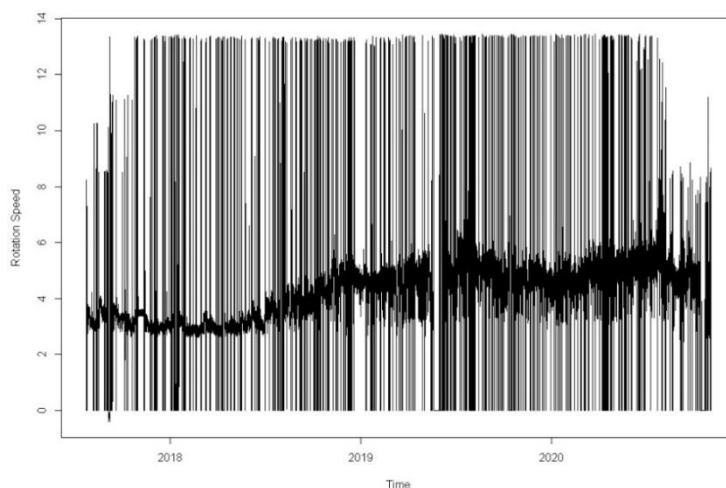


Figura 3-2 – Visualização gráfica dos registos válidos referentes à velocidade de rotação da prensa.

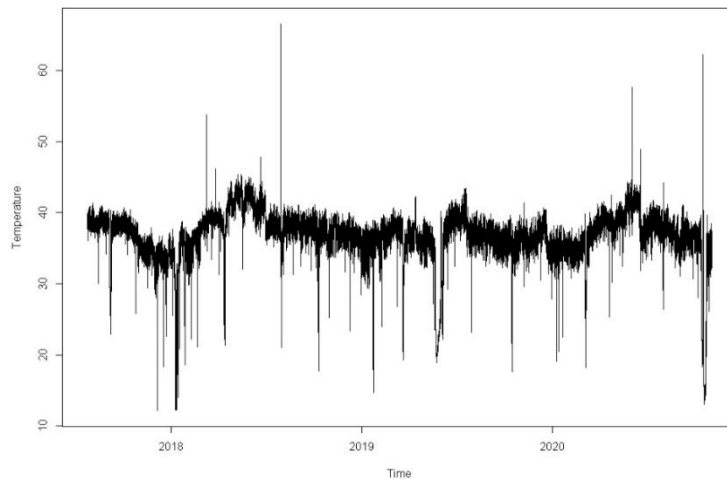


Figura 3-3 – Visualização gráfica dos registos válidos referentes à temperatura.

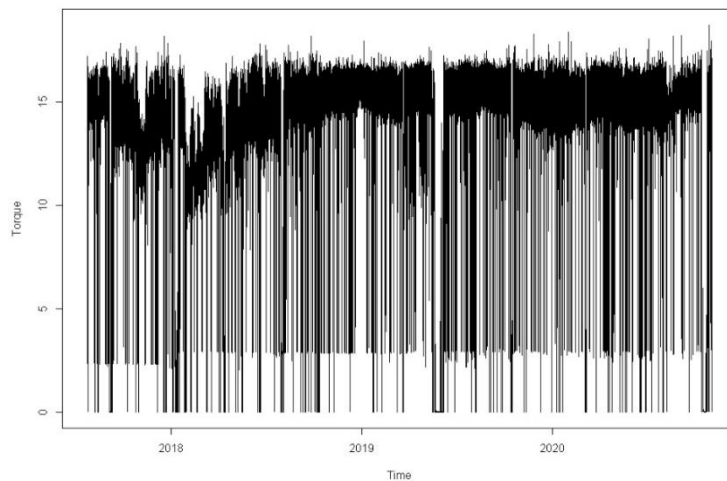


Figura 3-4 – Visualização gráfica dos registos válidos referentes ao torque.

Como referido anteriormente, no sentido de elaborar uma análise de sobrevivência, foi necessário executar alguns processos de tratamento de dados. O objetivo destes processos consistiu em obter uma tabela que permita associar os tempos de vida ao tipo de paragem que lhe sucedeu. Nesta fase é possível aplicar as funções existentes em livrarias específicas, para elaborar uma análise de sobrevivência. Estes procedimentos serão apresentados e discutidos no capítulo seguinte.

Apesar de, na sua generalidade, se tratar de processos não muito complexos, o modo como os dados se encontravam organizados inicialmente, bem como o volume de informação, conferiu um desafio adicional no desenvolvimento deste trabalho que, em parte, poderá não ser perceptível no resultado final.

3.3 Caracterização estatística

O tratamento de dados realizado permitiu obter um conjunto de 521 tempos de vida de média 2.11 dias e desvio padrão 3.31 dias. O valor máximo obtido foi de 24.08 dias e o valor mínimo é de 0.000694 dias (equivalente a 1 minuto), coerente com a

frequência de leituras realizada pelo sensor. A Figura 3-5 corresponde ao histograma dos tempos de sobrevivência onde é possível observar a distribuição destes dados. Verifica-se uma maior incidência de observações para tempos de sobrevivência reduzidos.

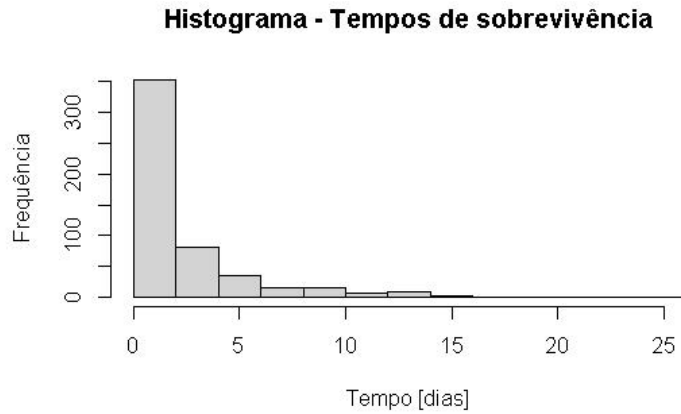


Figura 3-5 – Histograma dos tempos de sobrevivência.

Em estatística, os quartis dividem um conjunto de dados em quatro intervalos de igual probabilidade (25%). No conjunto de dados em análise, o primeiro quartil corresponde a 0.077778 (equivalente a 112 minutos); o segundo quartil, que corresponde à mediana e divide os dados em dois grupos com 50% de probabilidade de ocorrência, situa-se nos 0.735417 (aproximadamente 18 horas). Por fim, o terceiro quartil corresponde a 2.86 dias (aproximadamente 69 horas). Esta informação pode ser representada de forma visual através de um gráfico designado por *boxplot*, representado na Figura 3-6. Neste gráfico, o retângulo cinzento representa o intervalo entre o primeiro e o terceiro quartil. A linha vertical dentro do retângulo representa a mediana (equivalente ao segundo quartil). Os limites do *boxplot* são determinados com base no intervalo entre quartis, *IQR* que é dado pela diferença entre o terceiro e o primeiro quartil. Desta forma, os limites inferior e superior do *boxplot* são calculados por:

$$\text{Limite inferior} = Q_1 - 1.5 * IQR \quad (3.1)$$

$$\text{Limite superior} = Q_3 + 1.5 * IQR \quad (3.2)$$

Os valores que se encontram fora dos limites inferior e superior do *boxplot* são considerados *outliers* e estão representados por circunferências.

Tempos de vida

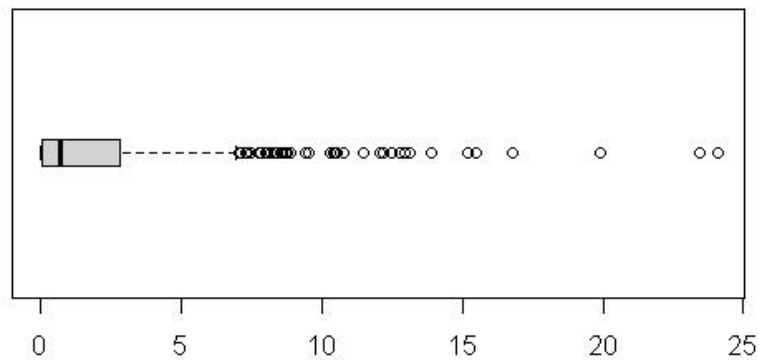


Figura 3-6 – *Boxplot* referente aos dados obtidos para os tempos de vida.

3.4 Resultados

3.4.1 Curvas de sobrevivência e de risco acumulado

Numa abordagem inicial, com recurso ao *software RStudio*, construíram-se as curvas de sobrevivência (Figura 3-7) e de risco acumulado (Figura 3-8) com base nos tempos de vida obtidos.

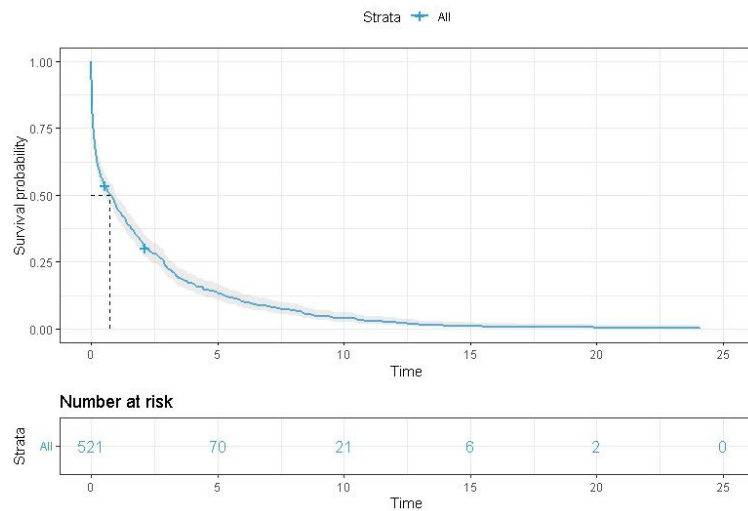


Figura 3-7 – Curva de sobrevivência com base nos tempos de vida identificados.

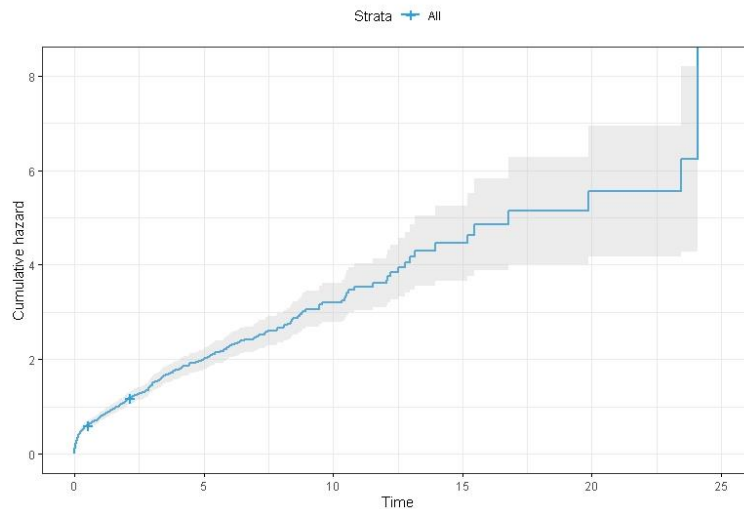


Figura 3-8 – Curva de risco acumulado com base nos tempos de vida identificados.

Analisando a curva de sobrevivência (Figura 3-7), é desde logo possível observar que foram identificados um total de 521 períodos de funcionamento. Por outro lado, apenas 70 tempos de vida são superiores a 5 dias e nenhum dos tempos de vida ultrapassa 25 dias. Desta forma, é possível concluir que a probabilidade de sobrevivência decai para metade muito rapidamente. Da mesma forma, a probabilidade de não se verificar falha nos primeiros cinco dias de funcionamento é de, aproximadamente, 12.5%.

A função de risco acumulado (Figura 3-8) representa, aproximadamente, o inverso da função de sobrevivência. Quer isto dizer que, no instante inicial do estudo, a probabilidade de sobrevivência é de 100% e o risco de falha é nulo. Ao longo do tempo, a esperança de vida (ou a probabilidade de não se verificar falha) irá necessariamente reduzir, pelo que o risco de falha aumenta progressivamente, atingindo o valor de 100% próximo dos 25 dias.

3.4.2 Análise de covariáveis

Como referido no capítulo anterior, recorreu-se aos dados disponíveis de alguns sensores instalados no equipamento para avaliar a influência destes parâmetros no tempo de sobrevivência da prensa. Tendo em conta o estudo que se apresenta, bem como os o tratamento de dados realizado, considerou-se adequado analisar os registos destas variáveis no instante imediatamente anterior ao instante da paragem. As Figuras 3-9 a 3-11 apresentam os gráficos relativos às variáveis em estudo, registados nos instantes anteriores a cada paragem. Note-se que, em cada instante, o valor da confiança nem sempre é comum a todos os sensores. Por outras palavras, existem instantes em que a confiança da leitura de intensidade de corrente é superior ao limite mínimo estabelecido, mas o mesmo não se verifica nas outras variáveis. Por este motivo e de forma idêntica ao descrito anteriormente, apenas foram considerados os instantes cujas leituras das variáveis em análise apresentam confiança superior ao limite estabelecido.

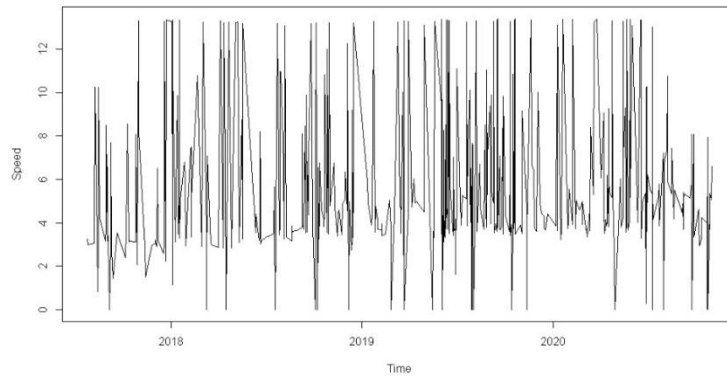


Figura 3-9 – Visualização gráfica dos registos referentes à velocidade no instante anterior a cada paragem.

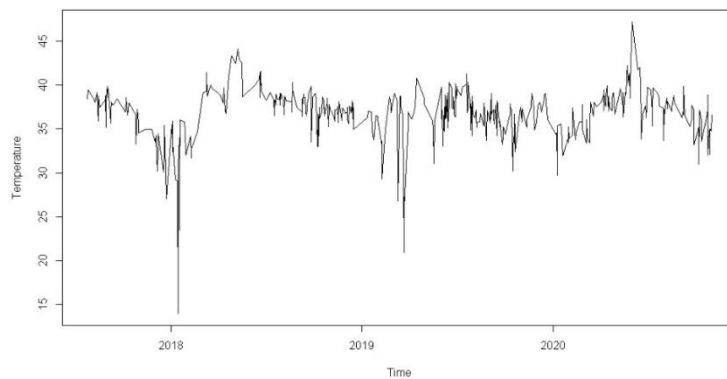


Figura 3-10 – Visualização gráfica dos registos referentes à temperatura no instante anterior a cada paragem.

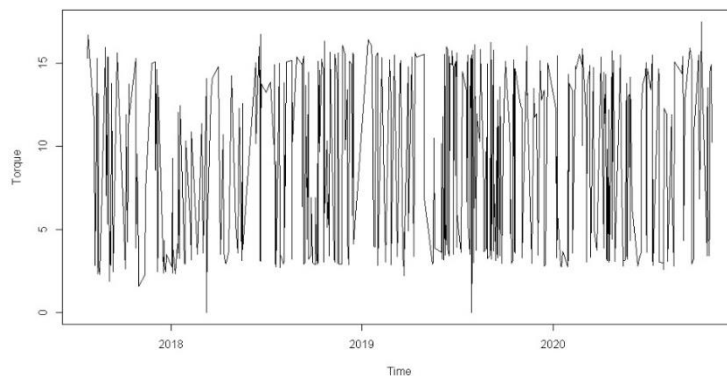


Figura 3-11 – Visualização gráfica dos registos referentes ao torque no instante anterior a cada paragem.

Com o objetivo de avaliar se a variação destes parâmetros tem influência nos tempos de vida do equipamento, definiu-se um intervalo de maior incidência para cada sensor. O intervalo foi definido com base nos gráficos das Figuras 3-2 a 3-4 e identificaram-se as leituras pertencentes ao intervalo, bem como as leituras inferiores e superiores, para cada variável em estudo (Tabela 3.9).

Manutenção Preditiva de uma Prensa de Secagem

Tabela 3.9 – Registo de leituras das variáveis em função do intervalo de maior incidência.

Variável	Intervalo	Leituras		
		Inferiores	Pertencentes	Superiores
Velocidade rotacional	[3; 7]	54	342	125
Temperatura	[30; 40]	18	468	35
Torque	[12; 16]	313	197	10

Com base nestes intervalos e com recurso ao *software RStudio*, construíram-se as curvas de sobrevivência baseadas no estimador *KM* para cada uma das variáveis em estudo. As Figuras 3-12 a 3-17 representam as curvas de sobrevivência e de risco acumulado das três variáveis em análise (ordenadas na seguinte sequência: velocidade de rotação, temperatura e torque).

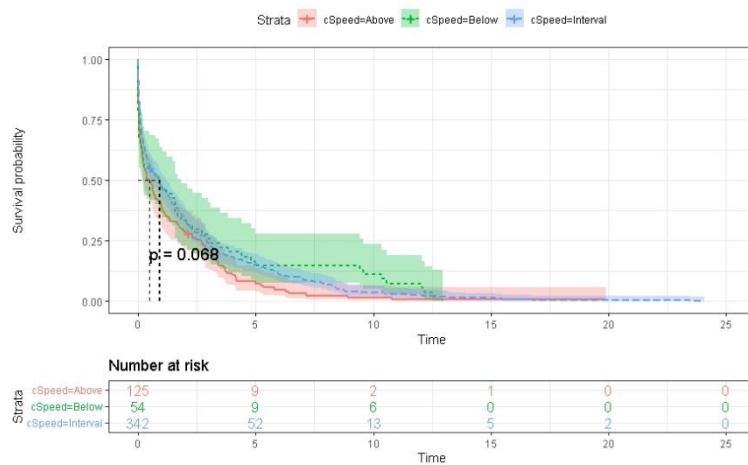


Figura 3-12 – Curva de sobrevivência com base na velocidade de rotação.

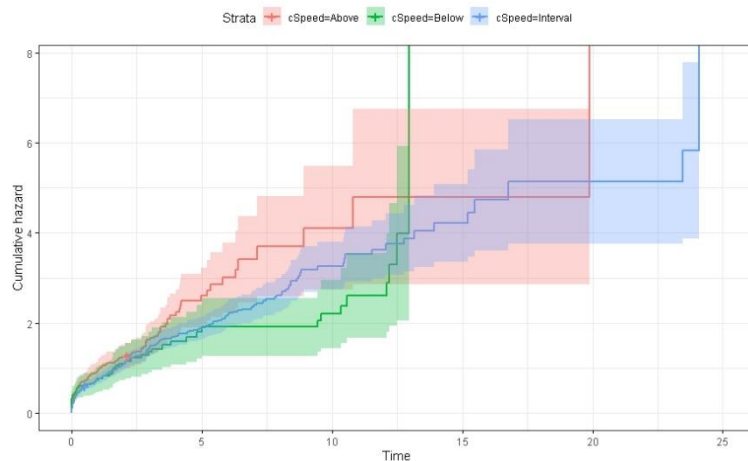


Figura 3-13 – Curva de risco acumulado com base na velocidade de rotação.

Nas curvas de sobrevivência referentes à velocidade de rotação (Figura 3-12), é possível observar que as três curvas, correspondentes a cada um dos grupos de observação definidos, apresentam um desenvolvimento similar. Apesar de a curva correspondente às observações de velocidade superiores ao intervalo definido apresentar uma probabilidade de sobrevivência inferior às restantes, a diferença

parece ser reduzida pelo que se torna difícil de concluir se a variação deste parâmetro tem influência significativa no comportamento do equipamento. De igual forma, apesar de as curvas de risco acumulado (Figura 3-13) parecerem apresentar um desenvolvimento mais distante, também não é possível concluir sobre o impacto deste parâmetro. Em ambos os gráficos, a mancha que envolve cada uma das linhas corresponde ao intervalo de confiança. Note-se que, no caso do gráfico de risco acumulado, este intervalo é mais significativo, pelo que é mais difícil concluir sobre a proximidade das curvas de risco.

No caso da variável correspondente à temperatura é possível observar um comportamento similar entre as curvas de sobrevivência (Figura 3-14) correspondentes às observações pertencentes ao intervalo e a curva das observações acima do intervalo. Por outro lado, a curva correspondente aos registos inferiores apresenta um intervalo de confiança maior e um comportamento mais diferenciado. Este aspeto está relacionado com o facto de se ter verificado um número reduzido de registos neste grupo. Por outro lado, é possível observar que a probabilidade de sobrevivência para valores de temperatura inferiores ao intervalo definido é aproximadamente metade do intervalo de referência. A observação das curvas de risco acumulado (Figura 3-15) conduzem a conclusões idênticas às anteriores. Contudo, as diferentes curvas correspondentes aos registos superiores e inferiores ao intervalo apresentam um comportamento em escada mais acentuado, devido ao menor número de registos.

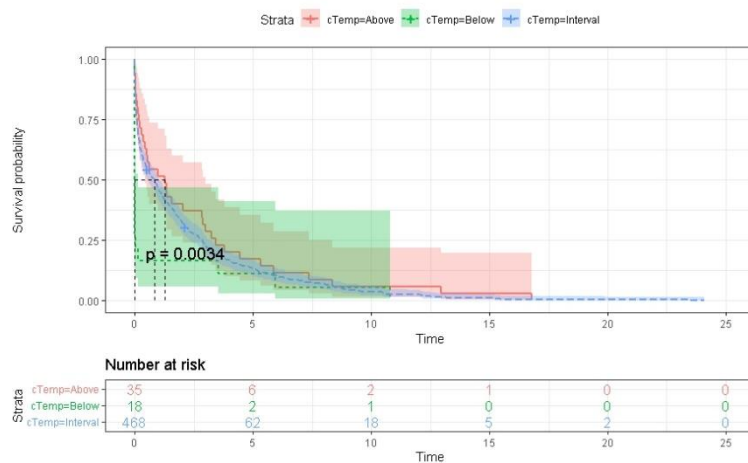


Figura 3-14 – Curva de sobrevivência com base na temperatura.

Manutenção Preditiva de uma Prensa de Secagem

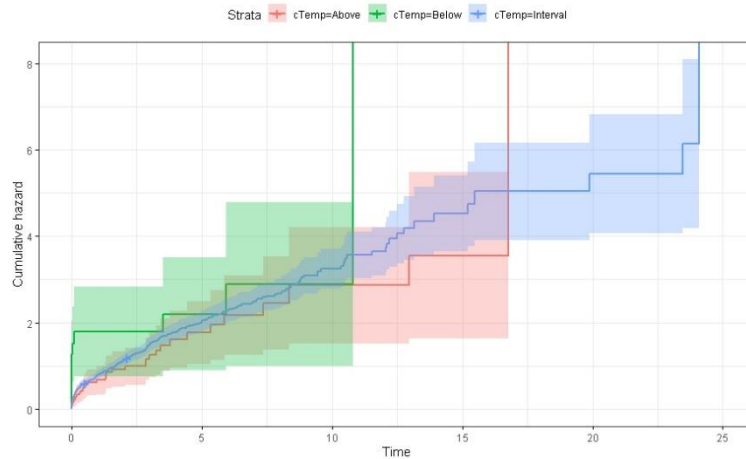


Figura 3-15 – Curva de risco acumulado com base na temperatura.

De forma idêntica ao caso anterior, a variável correspondente ao torque apresenta um número de registos acima do intervalo de referência relativamente reduzido em relação aos dois outros grupos. Por este motivo, o desenvolvimento da curva de sobrevivência apresenta um comportamento em escada mais pronunciado do que as curvas correspondentes a registos inferiores e dentro do intervalo (Figura 3-16). Por outro lado, uma vez que estes dois grupos apresentam um número de registos mais significativo, o seu desenvolvimento em escada é menos evidente e, conseqüentemente, os intervalos de confiança são mais reduzidos. A observação do gráfico permite observar um ligeiro desfasamento entre estas duas curvas, tratando-se do primeiro caso em que as manchas relativas aos intervalos de confiança não se sobrepõem. Esta situação repete-se quando se analisam as curvas de risco acumulado desta variável, pelo que conduz a uma perceção de que poderá existir evidência de que o nível de torque possa ter influência na probabilidade de sobrevivência (redução do risco de falha) da máquina.

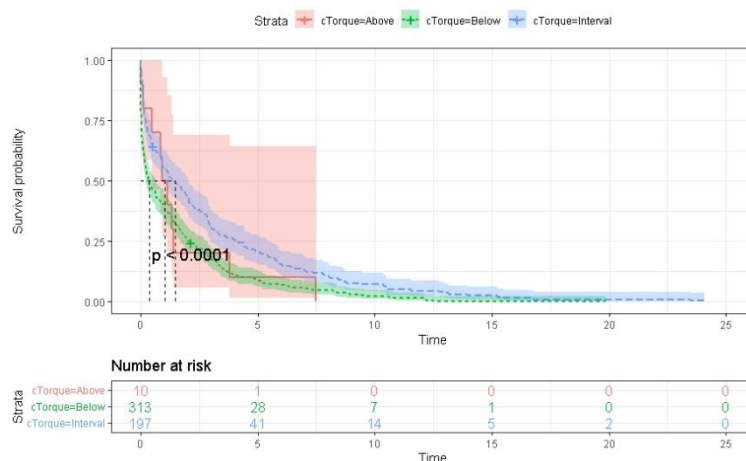


Figura 3-16 – Curva de sobrevivência com base o torque.

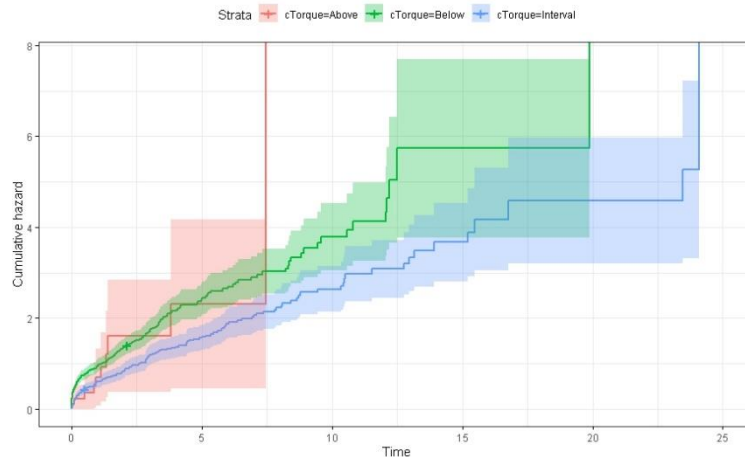


Figura 3-17 – Curva de risco acumulado com base no torque.

3.4.3 Comparação das curvas de sobrevivência

As curvas de sobrevivência e risco acumulado apresentadas na secção anterior são ferramentas importantes em análise de sobrevivência. A sua característica gráfica permite uma leitura mais rápida dos resultados, facilitando a sua interpretação. Contudo, a análise dos gráficos descrita anteriormente representa uma interpretação pouco rigorosa, podendo conduzir a conclusões imprecisas ou até grosseiras. Existem metodologias estatísticas que permitem contornar estas questões, no sentido de tirar conclusões rigorosas com base nos dados disponíveis. O teste de *Log-Rank* é um exemplo deste tipo de ferramentas. Esta metodologia consiste num teste não paramétrico sem pressupostos sobre a distribuição da função de sobrevivência, que tem como objetivo avaliar se existe evidência estatística que comprove a diferença entre curvas de sobrevivência (Kishore *et al.*, 2010).

Testes de hipóteses são ferramentas estatísticas muito utilizadas para avaliar uma população com base numa amostra. O procedimento inicia-se com a definição de uma hipótese (hipótese nula) para a população em estudo e o objetivo consiste em determinar se existe evidência que permita rejeitar essa hipótese (Ille & Milic, 2008). Os testes de hipóteses podem agrupar-se em paramétricos ou não paramétricos. Em geral, os testes paramétricos apresentam melhores resultados. No entanto, os testes paramétricos assumem que os parâmetros em análise seguem uma distribuição normal, o que nem sempre se verifica. Nestas situações, pode recorrer-se a testes não paramétricos (Chin & Lee, 2008). No caso de estudos de sobrevivência, devido à existência de dados censurados, é frequente recorrer-se a testes não paramétricos.

O teste de *Log-Rank* compara o número de eventos observados com o número de eventos esperado para cada grupo, caso a hipótese nula fosse verdadeira, ou seja, caso as curvas de sobrevivência fossem idênticas (Kishore *et al.*, 2010). Neste sentido, as hipóteses em análise podem ser definidas por:

$$H_0: S_{interval} = S_{Above} = S_{Below} \quad (3.3)$$

$$H_1: \text{Pelo menos uma das curvas é diferente.} \quad (3.4)$$

Por sua vez, a estatística de teste é definida por (LaMorte, 2016):

$$\chi^2 = \frac{[\sum O_{1t} - \sum E_{1t}]^2}{\sum Var(E_{1t})} \quad (3.5)$$

O numerador da equação (3.5) corresponde ao quadrado da diferença entre o somatório dos eventos (falhas) observadas ($\sum O_{it}$) e o somatório do número de eventos esperado ($\sum E_{it}$). O denominador corresponde à variância dos eventos esperados em cada intervalo de tempo e é calculada por (LaMorte, 2016):

$$Var(E_{1t}) = \frac{N_{1t} * N_{2t} * (N_t - O_t)}{N_t^2 * (N_t - 1)} \quad (3.6)$$

A Tabela 3.10 apresenta os resultados obtidos para a aplicação do teste de *Log-Rank* aos modelos de sobrevivência descritos nos capítulos anteriores, correspondentes a cada uma das variáveis em estudo.

Tabela 3.10 – Resultados do teste de *Log-Rank*.

Var.	Grupo	N.º Obs.	Falhas obs. (O)	Falhas esperadas (E)	$\frac{(O - E)^2}{E}$	$\frac{(O - E)^2}{V}$	χ^2	p-valor
Speed	Above	125	124	103.2	4.179	5.287	5.4	0.070
	Below	54	54	58.9	0.404	0.465		
	Interval	342	341	356.9	0.708	2.282		
Temp.	Above	35	35	41.36	0.9774	1.068	11.4	0.003
	Below	18	18	8.59	10.3009	10.548		
	Interval	468	466	469.05	0.0198	0.207		
Torque	Above	10	10	9.51	0.0257	0.0262	26.0	2.0E-06
	Below	313	312	255.61	12.4408	25.2185		
	Interval	197	196	252.89	12.796	25.7503		

A análise dos resultados do teste de *Log-Rank* pode ser realizada através do *p-valor*. Este indicador pode ser visto como a probabilidade de uma observação selecionada de forma aleatória seguir uma curva de sobrevivência diferente daquela que foi observada. Uma forma de avaliar o resultado do teste baseia-se na análise do *p-valor*. Se o *p-valor* for inferior ao limite estipulado (denominado por nível de significância e habitualmente assumindo o valor de 5%) poder-se-á rejeitar a hipótese nula. Por outras palavras, existe evidência estatística para concluir que pelo menos uma das curvas é suficientemente diferente de, pelo menos, uma das restantes curvas em estudo (Motulsky, 2023).

Observando a Tabela 3.10 é possível concluir que a hipótese nula é rejeitada no caso das variáveis referentes à temperatura ($p\text{-valor} = 0.03 < 0.05$) e ao torque ($p\text{-valor} = 2.0E-06 < 0.05$). Por outras palavras, poderá concluir-se que, pelo menos, uma das curvas de sobrevivência é suficientemente diferente de, pelo menos, uma das restantes. Neste sentido, poderá afirmar-se que estes fatores têm influência no comportamento do equipamento. Por outro lado, no caso da variável relativa à velocidade, obteve-se um $p\text{-valor}$ de 0.07, superior ao nível de significância de 0.05. Neste caso, não se pode rejeitar a hipótese nula, ou seja, as curvas não são suficientemente diferentes para concluir que este parâmetro tenha influência no comportamento da máquina. Contudo, importa referir que o $p\text{-valor}$ obtido é relativamente próximo do nível de significância assumido. Como referido anteriormente, poderá concluir-se que a probabilidade de obter um registo, selecionado de forma aleatória, que segue uma curva diferente da observada é de 7%. Note-se, ainda, que também é frequente considerar-se um nível de significância de 0.1. Neste caso, a hipótese nula seria rejeitada em todas as variáveis em análise. Estes resultados estão alinhados com a análise dos gráficos descrita no capítulo anterior e, de forma geral, confirmam que as variações dos parâmetros analisados têm influência no comportamento da prensa. Esta conclusão representa uma informação a ter em conta tanto na definição dos planos de manutenção como na monitorização do equipamento durante o seu funcionamento.

3.4.4 Modelos de regressão

Como referido anteriormente, a função de sobrevivência poderá envolver covariáveis (também designadas por variáveis explicativas) que poderão estar relacionadas com o tempo de sobrevivência. Na generalidade dos casos, este tipo de relações pode ser analisado através de modelos de regressão. Existem diversos tipos de modelos de regressão, tais como regressão linear ou não linear, entre outros. Este tipo de modelos baseia-se fortemente em distribuições normais. No entanto, estas distribuições não se revelam particularmente adequadas para descrever dados utilizados em análise de sobrevivência, que descrevem tempos de vida. Apesar de existirem distribuições de dados mais adequadas para este tipo de estudos, todos os modelos são considerados paramétricos e assumem pressupostos sobre a forma da função de risco (Motulsky, 2023).

O modelo de regressão de *Cox* é considerado um modelo semiparamétrico uma vez que não assume pressupostos quanto à distribuição dos dados de sobrevivência. Contudo, este modelo assume os seguintes pressupostos (LaMorte, 2016):

- Independência entre os tempos de vida de diferentes indivíduos;
- Proporcionalidade entre o estimador e o risco;
- A taxa de risco é constante ao longo do tempo.

Desta forma, o modelo de regressão de *Cox* pretende estimar o risco em função do tempo quando considerada a contribuição das covariáveis e pode ser formulado por (LaMorte, 2016):

$$h(t) = h_0(t) \exp (b_1X_1 + b_2X_2 + \dots b_nX_n) \quad (3.7)$$

A componente $h_0(t)$ representa a variação da função de risco no tempo de sobrevivência quando as covariáveis são nulas. A segunda componente consiste no exponencial da combinação linear das covariáveis, pelo que estas têm um impacto proporcional na função de risco. Os coeficientes do modelo de *Cox* representam a alteração no logaritmo da taxa de risco por unidade de variação da variável em causa, quando as restantes variáveis se mantêm constantes. Neste sentido, quando a exponencial do coeficiente é próxima da unidade, poderá concluir-se que a variável em causa não influencia o valor do risco. Por outro lado, quando a exponencial do coeficiente é inferior à unidade, a variável em causa está associada a uma redução do risco. Por outras palavras, o aumento dessa variável tem o efeito de prolongar o tempo de vida. De igual forma, quando a exponencial do coeficiente é superior à unidade, a respetiva variável está associada ao aumento do risco, ou seja, à redução do tempo de vida (LaMorte, 2016). A mesma análise pode ser realizada através do valor obtido para o próprio coeficiente. Coeficientes negativos correspondem a redução de risco e vice-versa.

A Tabela 3.11 apresenta os resultados obtidos para a aplicação do modelo de *Cox* no presente caso de estudo.

Tabela 3.11 – Modelo de regressão de *Cox*.

Var.	Grupo	coef	exp(coef)	p-valor
Speed	Above	-0.100	0.905	0.414
	Below	-0.265	0.767	0.088
Temp.	Above	-0.097	0.908	0.585
	Below	0.606	1.833	0.016
Torque	Above	0.321	1.378	0.324
	Below	0.522	1.685	1.33E-06

No caso de estudo que se apresenta, visto tratar-se de covariáveis com três níveis, a análise é realizada para cada covariável isoladamente, considerando que as restantes variáveis se mantêm constantes. Neste sentido, para cada uma das variáveis, o grupo de observações pertencentes ao intervalo de referência (*Interval*) foi, também nesta análise, considerado o grupo de referência.

Observando os resultados obtidos para a variável relativa à velocidade rotacional poderá concluir-se que o risco associado ao grupo de observações inferiores ao intervalo de referência (*Below*) apresenta um risco de falha de, aproximadamente, 23% inferior ao grupo de referência. No mesmo sentido, o grupo de observações superiores ao intervalo de referência (*Above*) apresenta, também, um risco de falha de, aproximadamente, 9.5% inferior ao grupo de referência. Estes resultados

sugerem que tanto velocidades superiores como inferiores ao grupo de referência apresentam maiores tempos de vida. Contudo, há outros aspetos que devem ser tidos em conta nesta análise.

Efetivamente, a construção do modelo de *Cox* está associada à realização de testes de hipóteses para cada variável, no sentido avaliar se existe evidência estatística de correlação entre cada variável e o tempo de vida (LaMorte, 2016). As hipóteses em estudo podem ser formuladas por:

$$H_0: \beta_j = 0 \quad (3.8)$$

$$H_1: \beta_j \neq 0 \quad (3.9)$$

Neste sentido, analisando os resultados obtidos para a variável relativa à velocidade de rotação é possível observar que o grupo superior apresenta um *p-valor* de 0.414, superior ao nível de significância de 0.05. Este resultado conduz à não rejeição da hipótese nula, pelo que se poderá concluir que β_j é nulo. Por outras palavras, este grupo de observações terá pouca expressividade na explicação da variável resposta do modelo de *Cox*. No caso das observações de velocidade inferiores ao intervalo de referência o *p-valor* obtido foi de 0.09, o que também conduz à não rejeição da hipótese nula. Contudo importa assinalar tratar-se de um valor muito mais reduzido do que no caso anterior, pelo que se poderá concluir que este grupo terá maior significado na explicação da variável resposta do modelo.

A análise das outras variáveis revela-se semelhante. No caso da temperatura, a exponencial do coeficiente sugere que temperaturas superiores apresentam uma redução do risco de falha, contudo o *p-valor* obtido é elevado, conduzindo à não rejeição da hipótese nula. Por outro lado, o grupo de observações inferiores ao intervalo de referência é indicador de um aumento do risco em 83%. Neste caso, o *p-valor* é de 0.016, inferior ao nível de significância, o que conduz à rejeição da hipótese nula. Assim, poderá concluir-se que este grupo de observações tem um impacto significativo na explicação do risco de falha do equipamento e que temperaturas inferiores ao grupo de referência apresentam menores tempos de sobrevivência.

No caso do torque, tanto o nível superior como o inferior apresentam maior risco de falha, respetivamente, 38% e 69% superior ao grupo de referência. No entanto, o *p-valor* do grupo de observações superiores (0.324) conduz à não rejeição da hipótese nula. Em sentido oposto, o *p-valor* do grupo de observações inferiores (1.33E-6) é muito inferior ao nível de significância, o que conduz à rejeição da hipótese nula e revela que este grupo de observações tem um peso significativo na explicação da variável resposta do modelo de *Cox*, correspondente ao risco de falha.

4 CONCLUSÕES E PROPOSTAS FUTURAS

O presente projeto teve como principal objetivo estudar a aplicabilidade de ferramentas estatísticas de fácil aplicação (quando comparadas com outro tipo de ferramentas, como as que se relacionam com inteligência artificial) ao contexto industrial, no sentido de retirar informação relevante para a tomada de decisão. Mais especificamente, este projeto incidiu na realização de uma análise de sobrevivência de uma prensa de secagem, inserida numa linha de produção de pasta de papel. Este tipo de análise tem como objetivo analisar o tempo de funcionamento habitual de um determinado equipamento, com base em curvas de sobrevivência e de risco acumulado. Adicionalmente, foram aplicadas ferramentas que permitem avaliar a influência de outros fatores no tempo de funcionamento do equipamento. Este tipo de análise permite retirar informação relevante, tanto para o desenvolvimento de planos de manutenção como no acompanhamento do funcionamento do equipamento.

O trabalho realizado teve como ponto de partida um conjunto de dados, registados ao longo de vários anos, provenientes de diversos sensores instalados no equipamento. Após a análise dos dados disponíveis, foram realizadas algumas transformações que permitiram determinar os tempos de funcionamento do equipamento. Numa primeira análise, foi possível concluir que o equipamento está, habitualmente, em funcionamento por períodos inferiores a 5 dias, nunca tendo ultrapassado os 25 dias consecutivos. Esta informação torna-se relevante na definição dos planos de manutenção, uma vez que permite, por exemplo, adequar os intervalos de paragem para ações de manutenção, em função do tempo expectável de funcionamento do equipamento.

O presente projeto contemplou, também, a análise de covariáveis, no sentido de avaliar a sua influência no tempo de funcionamento do equipamento. Foram analisadas as variáveis correspondentes à velocidade de rotação, temperatura e torque. Para cada uma das variáveis em estudo foram criados três intervalos de incidência de registos, para os quais foram construídas as respetivas curvas de sobrevivência e risco acumulado. Em paralelo com a construção das curvas foi realizado o teste de *Log-Rank* para cada uma das variáveis. As curvas de sobrevivência e de risco acumulado construídas com base nos intervalos definidos constituem uma ferramenta de fácil interpretação que, em conjunto com os testes de *Log-Rank*, permitem tirar conclusões quanto ao impacto destas variáveis no tempo de funcionamento do equipamento. Efetivamente o teste de *Log-Rank* demonstra que, para um nível de significância de 5%, diferentes leituras nos sensores correspondentes às covariáveis de temperatura e torque têm influência no tempo de funcionamento do equipamento. Apesar deste teste não identificar quais as curvas que se diferenciam, a observação das curvas de sobrevivência, poderá, por vezes, constituir um bom indicador nesta tarefa. Se, por um lado, no caso da temperatura esta informação não é evidente, por outro lado, as curvas de sobrevivência dos

diferentes intervalos de incidência relativos ao torque são indicadoras de que valores inferiores ao intervalo de referência aparentam conduzir a uma diminuição do tempo de sobrevivência.

Adicionalmente, a construção de modelos de regressão de *Cox* complementa estes resultados. No caso da covariável relativa à velocidade de rotação, os dados conduzem à conclusão de que, para um nível de significância de 5%, esta tem pouca influência na explicação da variável resposta. Por outro lado, os grupos de registos relativos a temperatura e torque inferiores aos respetivos intervalos de referência são indicadores de um aumento significativo do risco de falha, ou seja, redução do tempo de funcionamento. Estes resultados estão associados a *p-valor* inferiores ao nível de significância de 5%, pelo que se conclui que os dados comprovam que estas covariáveis têm impacto na variável resposta do modelo de *Cox*.

O conhecimento dos tempos de vida expectáveis de um determinado equipamento, bem como o modo como responde a variações das suas condições de funcionamento (avaliadas através da sua monitorização sob a forma de covariáveis), representa uma mais-valia para a gestão de manutenção, permitindo otimizar o seu funcionamento.

Em conclusão, o presente projeto constituiu um estudo inicial que comprova que a utilização deste tipo de ferramentas, mais especificamente a análise de sobrevivência, constitui uma abordagem válida na análise do funcionamento de equipamentos em contexto industrial, uma vez que permite tirar conclusões relevantes para o planeamento e monitorização da produção. O desenvolvimento de uma ferramenta automatizada, baseada em análise de sobrevivência, poderá constituir uma opção acessível de um ponto de vista financeiro, capaz de responder às necessidades de muitas organizações. Esta solução poderá, também, representar um ponto de partida numa transição sustentada para ferramentas mais complexas, tais como as que se baseiam em inteligência artificial.

Para trabalhos futuros, o desenvolvimento de uma ferramenta automatizada de tratamento de dados seria o próximo passo, no sentido de facilitar a inclusão de registos de equipamentos idênticos, pertencentes a outras linhas de produção. A monitorização e avaliação do impacto de outras covariáveis constitui outro aspeto que poderá providenciar informação relevante sobre o comportamento deste tipo de equipamentos. Por fim, a adição de mais registos ao conjunto de dados em análise, tanto por via da monitorização continuada deste equipamento como pela inclusão de outros, poderá permitir obter mais informação relevante para a melhoria do funcionamento deste tipo de máquinas, permitindo otimizar os momentos de paragem, de forma a evitar falhas graves.

REFERÊNCIAS BIBLIOGRÁFICAS

- Araújo, C. (2010). *Optimização do processo de lavagem de pasta branqueada de eucalipto* [Universidade de Aveiro].
<https://ria.ua.pt/bitstream/10773/3157/1/2010001398.pdf>
- Borges, A. I. M. (2014). *Análise de Sobrevivência com o R*.
<https://digituma.uma.pt/handle/10400.13/732>
- Chin, R., & Lee, B. Y. (2008). *Principles and Practice of Clinical Trial Medicine*. Elsevier.
<https://doi.org/10.1016/B978-0-12-373695-6.X0001-4>
- DATAtab Team. (2023). *Kaplan Meier Curve*. Retrieved June, 2023 from
<https://Datatab.Net/Tutorial/Kaplan-Meier-Curve>.
- Ille, T., & Milic, N. (2008). Statistical Tests. In W. Kirch (Ed.), *Encyclopedia of Public Health* (pp. 1341–1344). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5614-7_3349
- IPQ. (2007). *NP-EN-13306 Terminologia da manutenção*.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457.
<https://doi.org/10.2307/2281868>
- Kishore, J., Goel, M., & Khanna, P. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, 1(4), 274.
<https://doi.org/10.4103/0974-7788.76794>
- LaMorte, W. W. (2016, June 3). *Comparing Survival Curves*. Retrieved June, 2023 from
https://Sphweb.Bumc.Bu.Edu/Otl/Mph-Modules/Bs/Bs704_survival/BS704_Survival5.html.
- Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., & Cardoso, A. M. (2021). Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies*, 14(21).
<https://doi.org/10.3390/en14216958>
- Motulsky, H. J. (2023). *GraphPad Statistics Guide*. Retrieved June, 2023 from
<http://www.Graphpad.Com/Guides/Prism/10/Statistics/Index.htm>.
- Ramalho, J. (2019). *Monitorização de Equipamentos de Baixa Rotação Engenharia Mecânica Orientador(es): Prof. Virgínia Isabel Monteiro Nabais Infante* [Instituto Superior Técnico].
https://fenix.tecnico.ulisboa.pt/downloadFile/563345090418781/Tese_de_Mestrado_78511.pdf
- Ramos, P. L. (2018). *Bayesian and classical inference for the generalized gamma distribution and related models* [Universidade de São Paulo].
<https://www.researchgate.net/publication/340092206>

Singh, V. (2023, August 21). *PDF vs. CDF: Difference Between PDF and CDF*. Retrieved September, 2023 from <https://www.shiksha.com/online-courses/articles/difference-between-pdf-and-cdf/>

Xia, L., Zheng, P., Li, X., Gao, R. X., & Wang, L. (2022). Toward cognitive predictive maintenance: A survey of graph-based approaches. In *Journal of Manufacturing Systems* (Vol. 64, pp. 107–120). Elsevier B.V. <https://doi.org/10.1016/j.jmsy.2022.06.002>



**Instituto Superior
de Engenharia**

Politécnico de Coimbra