

**UNIVERSIDADE FEDERAL DE VIÇOSA**

**Análise de sobrevivência aplicada no melhoramento genético do cafeeiro**

Lucas Pereira Belo  
*Magister Scientiae*

**VIÇOSA - MINAS GERAIS  
2025**

**LUCAS PEREIRA BELO**

**Análise de sobrevivência aplicada no melhoramento genético do cafeeiro**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

Orientador: Paulo Cesar Emiliano

**VIÇOSA - MINAS GERAIS  
2025**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

B452a  
2025  
Belo, Lucas Pereira, 1998-  
Análise de sobrevivência aplicada no melhoramento  
genético do cafeeiro / Lucas Pereira Belo. – Viçosa, MG, 2025.  
1 dissertação eletrônica (127 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Paulo César Emiliano.

Dissertação (mestrado) - Universidade Federal de Viçosa,  
Departamento de Estatística, 2025.

Referências bibliográficas: f. 94-99.

DOI: <https://doi.org/10.47328/ufvbbt.2025.409>

Modo de acesso: World Wide Web.

1. Análise de sobrevivência (Biometria). 2. Censura. 3. Café  
- Melhoramento genético. 4. Café - Resistência a doenças e  
pragas. I. Emiliano, Paulo César, 1982-. II. Universidade Federal  
de Viçosa. Departamento de Estatística. Programa de  
Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 519.546

**LUCAS PEREIRA BELO**

**Análise de sobrevivência aplicada no melhoramento genético do cafeeiro**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 24 de fevereiro de 2025.

Assentimento:

---

Lucas Pereira Belo  
Autor

---

Paulo Cesar Emiliano  
Orientador

Essa dissertação foi assinada digitalmente pelo autor em 16/06/2025 às 15:43:47 e pelo orientador em 16/06/2025 às 16:31:51. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **Q7CB.2BAX.LKDV** e clique no botão 'Validar documento'.

## **AGRADECIMENTOS**

Agradeço, em primeiro lugar, a Deus, por me conceder diariamente forças e determinação para enfrentar os desafios e superar os obstáculos do caminho, e por colocar pessoas extraordinárias em minha vida. À toda minha família, meus pais, Sandra e Sivanil, minhas irmãs, Geanes e Geigiane, meus sobrinhos Kaylan e Eloisa, ao meu cunhado Marcelos, e a minha companheira Faissandra, pelo apoio incondicional em todas as etapas durante este processo. Sou profundamente grato pelas orações, pelo apoio constante, pelo ombro sempre acolhedor, por tornarem esta jornada mais leve e menos desgastante, e por estarem ao meu lado em todos os momentos, contribuindo de forma essencial para esta conquista que também é de vocês.

Ao Departamento de Estatística da Universidade Federal de Viçosa (DET - UFV), por me fornecer apoio e a estrutura necessária para realização do mestrado. Ao meu orientador, Prof. Dr. Paulo Cesar Emiliano, pelo apoio, dedicação, e por suas sugestões e revisões que foram de grande valia para a conclusão da dissertação. Agradeço também à Prof. Dra. Eveline Teixeira Caixeta pela disponibilização dos dados e a doutoranda Bruna Mariz pelo suporte e apoio nas revisões.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## RESUMO

BELO, Lucas Pereira, M.Sc., Universidade Federal de Viçosa, fevereiro de 2025. **Análise de sobrevivência aplicada no melhoramento genético do cafeeiro.** Orientador: Paulo Cesar Emiliano.

Este estudo aplicou técnicas de análise de sobrevivência a dados do genótipo resultante do cruzamento entre o Híbrido de Timor MG 0357 e o Tupi Amarelo IAC 5162, com foco na produtividade, resistência a doenças e qualidade da bebida, bem como na influência de covariáveis sobre a produção. O evento de interesse foi definido com base na produção média anual das testemunhas Paraíso MG H419-1 e Catuaí Vermelho IAC 144. Foram utilizados o estimador de Kaplan-Meier e os testes de log-rank e Wilcoxon para a comparação entre grupos, enquanto modelos de regressão de Cox, com aproximação da função de verossimilhança parcial, e modelos para dados agrupados avaliaram a relação entre as covariáveis e o tempo até a primeira produção. A adequação dos modelos foi verificada por meio dos resíduos de Schoenfeld e dos resíduos de Cox-Snell; utilizou-se ainda a curva ROC (*Receiver Operating Characteristic*) para a validação dos modelos para dados agrupados. Os resultados indicaram resistência do cafeeiro à ferrugem, bem como uma associação entre baixo grau de incidência das doenças e maior produtividade. Os testes de log-rank e Wilcoxon detectaram diferenças significativas entre os grupos relacionados a todas as covariáveis analisadas, com destaque para a covariável ciclo de maturação, na qual foram identificados diferentes ciclos produtivos com elevadas capacidades de produção. Os modelos de Cox, ajustados com aproximações da função de verossimilhança parcial, revelaram violações nas pressuposições de riscos proporcionais. Nos modelos discretos, as covariáveis altura da planta, vigor vegetativo, temperatura, precipitação e número de nós nos ramos plagiotrópicos mostraram-se significativas na relação com a produtividade. A curva ROC indicou um desempenho excelente desses modelos. A análise de sobrevivência demonstrou-se eficaz na identificação de fatores associados à produtividade do cafeeiro, destacando-se a aplicação das curvas de Kaplan-Meier e dos testes log-rank e Wilcoxon. Embora os modelos discretos impliquem certa perda de informação, mostraram-se eficientes para a identificação das covariáveis importantes para a produção.

Palavras-chave: empates; censura; dados agrupados

## ABSTRACT

BELO, Lucas Pereira, M.Sc., Universidade Federal de Viçosa, February, 2025. **Survival analysis applied to the genetic improvement of coffee plants.** Adviser: Paulo Cesar Emiliano.

This study applied survival analysis techniques to data from the genotype resulting from the cross between the Timor Hybrid MG 0357 and the Yellow Tupi IAC 5162, focusing on productivity, disease resistance and beverage quality, as well as the influence of covariates on production. The event of interest was defined based on the average annual production of the controls Paraíso MG H419-1 and Catuaí Vermelho IAC 144. The Kaplan-Meier estimator and the log-rank and Wilcoxon tests were used for comparison between groups, while Cox regression models, with approximation of the partial likelihood function, and models for grouped data evaluated the relationship between the covariates and the time until the first production. The adequacy of the models was verified by means of Schoenfeld residuals and Cox-Snell residuals; the ROC (Receiver Operating Characteristic) curve was also used to validate the models for grouped data. The results indicated resistance of coffee plants to rust, as well as an association between low incidence of diseases and higher productivity. The log-rank and Wilcoxon tests detected significant differences between the groups related to all the covariates analyzed, with emphasis on the covariate maturation cycle, in which different production cycles with high production capacities were identified. The Cox models, adjusted with approximations of the partial likelihood function, revealed violations in the assumptions of proportional hazards. In the discrete models, the covariates plant height, vegetative vigor, temperature, precipitation and number of nodes in plagiotropic branches were shown to be significant in the relationship with productivity. The ROC curve indicated excellent performance of these models. The survival analysis proved to be effective in identifying factors associated with coffee productivity, with emphasis on the application of the Kaplan-Meier curves and the log-rank and Wilcoxon tests. Although the discrete models imply a certain loss of information, they proved to be efficient in identifying the covariates important for production.

Keywords: ties; censorship; grouped data

## LISTA DE FIGURAS

Figura 1	– Mecanismo de censura tipo I. . . . .	16
Figura 2	– Mecanismo de censura tipo II. . . . .	17
Figura 3	– Mecanismo de censura tipo III. . . . .	17
Figura 4	– Sobrevivência e respectivos intervalos de 95% de confiança. . . . .	19
Figura 5	– Curvas de sobrevivência: Grupo 1 versus Grupo 2. . . . .	22
Figura 6	– Passo a passo da metodologia utilizada. . . . .	40
Figura 7	– Curva de produção na comparação com a testemunha Paraíso. . . . .	41
Figura 8	– Produtividade estimada via Kaplan-Meier para cada nível da covariável fileira: comparação com a testemunha Paraíso. . . . .	43
Figura 9	– Produtividade estimada via Kaplan-Meier para cada nível da covariável bloco: comparação com a testemunha Paraíso. . . . .	43
Figura 10	– Produtividade estimada via Kaplan-Meier para cada nível da covariável ferrugem: comparação com a testemunha Paraíso. . . . .	45
Figura 11	– Produtividade estimada via Kaplan-Meier para cada nível da covariável bicho mineiro. . . . .	46
Figura 12	– Produtividade estimada via Kaplan-Meier para cada nível da covariável cercosporiose: comparação com a testemunha Paraíso. . . . .	48
Figura 13	– Produtividade estimada via Kaplan-Meier para cada nível da covariável uniformidade de maturação dos frutos. . . . .	50
Figura 14	– Produtividade estimada via Kaplan-Meier para cada nível da covariável tamanho do fruto: comparação com a testemunha Paraíso. . . . .	52
Figura 15	– Produtividade estimada via Kaplan-Meier para cada nível da covariável ciclo de maturação: comparação com a testemunha Paraíso. . . . .	54
Figura 16	– Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação Breslow: Comparação com a testemunha Paraíso. . . . .	57
Figura 17	– Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação Efron: Comparação com a testemunha Paraíso. . . . .	60
Figura 18	– Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação exata: Comparação com a testemunha Paraíso. . . . .	63
Figura 19	– Curva ROC, modelo logístico. . . . .	66
Figura 20	– Produtividade estimada via Kaplan-Meier: comparação com a testemunha Catuaí Vermelho. . . . .	67
Figura 21	– Produtividade estimada via Kaplan-Meier para cada nível da covariável fileira: comparação com a testemunha Catuaí Vermelho. . . . .	69
Figura 22	– Produtividade estimada via Kaplan-Meier para cada nível da covariável bloco: comparação com a testemunha Catuaí Vermelho. . . . .	70
Figura 23	– Produtividade estimada via Kaplan-Meier para cada nível da covariável ferrugem: comparação com a testemunha Catuaí Vermelho. . . . .	72
Figura 24	– Produtividade estimada via Kaplan-Meier para cada nível da covariável bicho mineiro: comparação com a testemunha Catuaí Vermelho. . . . .	73

Figura 25 – Produtividade estimada via Kaplan-Meier para cada nível da covariável cercosporiose: comparação com a testemunha Catuaí Vermelho. .	75
Figura 26 – Produtividade estimada via Kaplan-Meier para cada nível da covariável uniformidade dos frutos: comparação com a testemunha Catuaí Vermelho. . . . .	77
Figura 27 – Produtividade estimada via Kaplan-Meier para cada nível da covariável tamanho do fruto: comparação com a testemunha Catuaí Vermelho.	78
Figura 28 – Produtividade estimada via Kaplan-Meier para cada nível da covariável ciclo de maturação: comparação com a testemunha Catuaí Vermelho.	80
Figura 29 – Análise dos resíduos de Cox-Snell do modelo de aproximação Breslow: Comparação com a testemunha Catuaí Vermelho. . . . .	83
Figura 30 – Análise dos resíduos de Cox-Snell do modelo de aproximação Efron: Comparação com a testemunha Catuaí Vermelho. . . . .	86
Figura 31 – Análise dos resíduos de Cox-Snell do modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho. . . . .	89
Figura 32 – Curva ROC, modelo de riscos proporcionais. . . . .	92

## LISTA DE TABELAS

Tabela 1	– Tabela de contingência associada ao tempo de falha $t_j$ . . . . .	23
Tabela 2	– Covariáveis utilizadas no estudo. . . . .	36
Tabela 3	– Covariáveis categóricas. . . . .	37
Tabela 4	– Covariáveis contínuas. . . . .	38
Tabela 5	– Produtividade estimada via Kaplan-Meier em comparação com a testemunha Paraíso. . . . .	41
Tabela 6	– Teste de Wilcoxon na comparação das curvas de produção entre os níveis da covariável bloco. . . . .	44
Tabela 7	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável ferrugem. . . . .	45
Tabela 8	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável bicho-mineiro. . . . .	46
Tabela 9	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável cercosporiose. . . . .	47
Tabela 10	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável uniformidade de maturação dos frutos. . . . .	49
Tabela 11	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável tamanho dos frutos. . . . .	51
Tabela 12	– Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável ciclo de maturação. . . . .	53
Tabela 13	– Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Breslow. . . . .	55
Tabela 14	– Teste da razão de verossimilhança para o modelo de aproximações de Breslow. . . . .	56
Tabela 15	– Análise dos coeficientes razão de riscos e intervalos de confiança para o modelo de aproximação Breslow. . . . .	56
Tabela 16	– Teste de Schoenfeld aplicado as covariáveis do modelo de aproximação de Breslow. . . . .	57
Tabela 17	– Seleção de covariáveis usando o modelo de regressão de Cox, com aproximação de Efron. . . . .	58
Tabela 18	– Teste da razão de verossimilhança para o modelo de aproximações de Efron. . . . .	59
Tabela 19	– Análise dos coeficientes razão de risco e intervalos de confiança, no modelo de Efron. . . . .	59
Tabela 20	– Teste de Schoenfeld aplicado às covariáveis do modelo com aproximação de Efron. . . . .	60
Tabela 21	– Seleção de covariáveis considerando o modelo de regressão de Cox, com aproximação exata. . . . .	61

Tabela 22	– Teste da razão de verossimilhança para o modelo de aproximações exata. . . . .	62
Tabela 23	– Análise dos coeficientes, razão de riscos e intervalos de confiança, no modelo de aproximação exata. . . . .	62
Tabela 24	– Teste de Schoenfeld aplicado as covariáveis do modelo com aproximação exata. . . . .	63
Tabela 25	– Resumo dos dados apresentados pela tabela de vida. . . . .	64
Tabela 26	– Estimativas dos parâmetros dos modelos finais do processo <i>stepwise</i> para os modelos de Cox e logístico. . . . .	64
Tabela 27	– Produtividade estimada via Kaplan-Meier em comparação com a testemunha Catuaí Vermelho. . . . .	67
Tabela 28	– Resultados dos testes de log-rank e Wilcoxon para comparação entre os níveis dos diferentes blocos. . . . .	70
Tabela 29	– Testes de log-rank e Wilcoxon na comparação das curvas de produção de café sob diferentes níveis da covariável ferrugem. . . . .	71
Tabela 30	– Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável bicho-mineiro. . . . .	73
Tabela 31	– Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável cercosporiose. . . . .	74
Tabela 32	– Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável uniformidade de maturação dos frutos. . . . .	76
Tabela 33	– Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável tamanho dos frutos. . . . .	78
Tabela 34	– Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável ciclo de maturação. . . . .	79
Tabela 35	– Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Breslow. . . . .	81
Tabela 36	– Teste da razão de verossimilhança para o modelo de aproximações de Breslow: Comparação com a testemunha Catuaí Vermelho. . . . .	82
Tabela 37	– Análise dos coeficientes, razão de risco e intervalos de confiança para o modelo Breslow: Comparação com a testemunha Catuaí Vermelho. . . . .	82
Tabela 38	– Teste de Schoenfeld aplicado às covariáveis do modelo de aproximação de Breslow: Comparação com a testemunha Catuaí Vermelho. . . . .	83
Tabela 39	– Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Efron: Comparação com a testemunha Catuaí Vermelho. . . . .	84
Tabela 40	– Teste da razão de verossimilhança para o modelo de aproximações de Efron: Comparação com a testemunha Catuaí Vermelho. . . . .	85
Tabela 41	– Análise dos coeficientes, razão de risco e intervalos de confiança para o modelo Efron: Comparação com a testemunha Catuaí Vermelho. . . . .	85
Tabela 42	– Teste de Schoenfeld aplicado as covariáveis do modelo de aproximação de Efron: Comparação com a testemunha Catuaí Vermelho. . . . .	86
Tabela 43	– Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação exata: Comparação com a testemunha Catuaí Vermelho. . . . .	87

Tabela 44 – Teste da razão de verossimilhança para o modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho. . . . .	88
Tabela 45 – Análise dos coeficientes, razão de risco e respectivos intervalos de confiança para o modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho. . . . .	88
Tabela 46 – Teste de Schoenfeld aplicado às covariáveis do modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho. . . . .	89
Tabela 47 – Resumo dos dados apresentados pela tabela de vida: Comparação com a testemunha Catuaí Vermelho. . . . .	90
Tabela 48 – Estimativas dos parâmetros dos modelos finais do processo <i>stepwise</i> para os modelos de Cox e logístico: Comparação com a testemunha Catuaí Vermelho. . . . .	91

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>15</b>
2.1	Análise de sobrevivência	15
2.2	Censura	15
2.3	Funções do tempo de sobrevivência	18
2.3.1	Função de sobrevivência	18
2.3.2	Função de risco	19
2.3.3	Função taxa de falha acumulada	20
2.3.4	Estimador de Kaplan-Meier (K-M)	20
2.3.5	Tabela de vida	21
2.4	Comparação de curvas de sobrevivência	21
2.4.1	Teste log-rank	22
2.4.2	Teste de Wilcoxon	24
2.5	Modelo de riscos proporcionais de Cox	24
2.6	Verossimilhança parcial	25
2.7	Aproximação da verossimilhança parcial	26
2.7.1	Aproximação exata	26
2.7.2	Aproximação de Efron	27
2.7.3	Aproximação de Breslow	27
2.8	Modelos de regressão discretos	28
2.8.1	Modelo de taxas de falhas proporcionais	29
2.8.2	Modelo logístico	30
2.9	Critérios de informação	30
2.9.1	Critério de informação de Akaike - AIC	31
2.9.2	Critério de informação Bayesiano - BIC	31
2.10	Metodologia de Collet	32
2.11	Avaliação dos modelos	32
2.11.1	Resíduos de Cox e Snell	33
2.11.2	Resíduos <i>martingal</i>	33
2.11.3	Resíduos Schoenfeld	33
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>35</b>
3.1	Descrição dos dados	35
3.2	Métodos	38
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>40</b>
4.1	Paraíso	40
4.1.1	Comparações das curvas de produção	42
4.1.2	Modelo final utilizando aproximação de Breslow	54

4.1.3	Adequação do modelo . . . . .	56
4.1.4	Modelo final utilizando aproximação de Efron . . . . .	58
4.1.5	Adequação do modelo . . . . .	59
4.1.6	Modelo final utilizando aproximação exact . . . . .	60
4.1.7	Adequação do modelo . . . . .	62
4.1.8	Modelos discretos . . . . .	63
4.1.9	Modelos finais . . . . .	64
4.2	Catuaí Vermelho . . . . .	67
4.2.1	Comparação das curvas de produção . . . . .	68
4.2.2	Modelo final utilizando aproximação de Breslow . . . . .	80
4.2.3	Adequação do modelo . . . . .	82
4.2.4	Modelo final utilizando aproximação de Efron . . . . .	83
4.2.5	Adequação do modelo . . . . .	85
4.2.6	Modelo final utilizando a aproximação exata . . . . .	86
4.2.7	Adequação do modelo . . . . .	88
4.2.8	Modelos finais . . . . .	90
<b>5</b>	<b>CONCLUSÕES . . . . .</b>	<b>93</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>94</b>
	<b>APÊNDICE . . . . .</b>	<b>100</b>

# 1 INTRODUÇÃO

O Brasil é uma potência na produção de café, o país produz anualmente milhões de sacas do produto. De acordo com dados fornecidos pela CONAB (2025) constata-se que a safra 2023-2024 total dos cafés do Brasil para o período foi de 54,79 milhões de sacas de 60 kg, as quais representam em torno de 34% da produção global. Além disso, o país é o maior produtor e o segundo maior consumidor de café do mundo, atrás apenas dos Estados Unidos (ABIC, 2023). Além da vasta produção de café, vale destacar a relevância do *Coffea arabica*, no cenário brasileiro, não apenas pelo volume produzido, mas também pela sua qualidade e variedade de sabores.

Originário da Etiópia, o *Coffea arabica* chegou ao Brasil pelas mãos do oficial Francisco de Mello Palheta, o qual trouxe mudas vindas da Guiana e desembarcadas no norte do Brasil. Com o tempo, o cultivar foi ganhando a região sudeste do Brasil, pela facilidade na exportação e também pelo fator climático (NAGAY, 1999). Ainda de acordo com Nagay (1999) no estado do Rio de Janeiro ocorreram as primeiras grandes plantações. Em 1860 o estado representava mais de 78% da produção nacional de café, com grande destaque da província de Vassouras que, na época, se tornou a capital do café no Brasil.

Com o crescimento da indústria cafeeira, novas regiões do Brasil foram exploradas para cultivo, culminando na proeminência atual de estados como Minas Gerais e São Paulo na produção nacional, processo que se deu pelo Vale do Paraíba (FILETTO E ALENCAR, 2001). Atualmente, Minas Gerais é o maior produtor nacional de café, com uma produção de aproximadamente 50% do que é produzido nacionalmente, sendo o estado o maior produtor desta qualidade do país. Somente em 2024, estima-se uma produção de 29 milhões de sacas em solo mineiro (FERREIRA e CAVATON, 2024).

Novos cultivares surgem frequentemente no mercado, oriundos de um longo processo de melhoramento através de melhoristas que buscam, em geral, uma alta resistência a pragas, produtividade, vigor, longevidade e plantas adaptadas às diferentes regiões do país, como destacam Caixeta *et al.* (2022). Diversas metodologias existem para auxiliar nesse processo, principalmente quando se fala em um contexto genético, diante disso, torna-se necessário incorporar novas tecnologias que possam incrementar a dinâmica e a capacidade de resposta dos fatores ligados ao fenótipo das plantas durante esse processo, de modo a corroborar com a análise genotípica.

Uma forma de analisar tais fatores é por meio de covariáveis ligadas as plantas ao longo do tempo, desde o plantio até o início da produção. Durante esse processo, é comum a ocorrência de replantas, que consistem na reposição de mudas para substituir plantas que morreram antes de atingir a fase produtiva, conforme indicado pela CNA (2017). De acordo com Duarte *et al.* (2011) as plantas que morreram devem ser removidas e substituídas por novas mudas, processo que, em geral, ocorre no segundo ano do plantio.

A análise de sobrevivência apresenta-se como uma técnica promissora, pois se caracteriza como uma metodologia longitudinal, na qual o interesse está na modelagem do tempo até a ocorrência de um evento específico. No contexto deste trabalho, o evento de interesse é o alcance de uma produção igual ou superior a produção média das testemunhas ao longo dos anos de acompanhamento.

Um diferencial dessa abordagem em relação a outras metodologias estatísticas é a inclusão de observações censuradas no estudo. Neste caso, consideram-se censuradas as plantas que

morreram antes de atingir o evento de interesse ou que não o alcançaram até o final do estudo. De acordo com Colosimo e Giolo (2006), ainda que incompletas, as observações censuradas são fundamentais para a análise, pois sua omissão pode levar a conclusões enviesadas.

A comparação entre diferentes curvas de sobrevivência “aqui interpretadas como curvas de produtividade” para distintos níveis de covariáveis categóricas, como a presença ou ausência de ferrugem, pode ser realizada por meio de testes estatísticos apropriados, como os de Wilcoxon e Log-rank. O teste de Wilcoxon mostra-se mais eficaz na detecção de diferenças iniciais na sobrevida (MARTINEZ e NARANJO, 2010). Por outro lado, de acordo com Jayawardene e Yang (1996), o teste de Log-Rank é mais poderoso para detectar diferenças em tempos de sobrevida posteriores.

Essa diversificação nas características dos testes permite que diferentes períodos da produtividade do cultivar sejam examinados ao longo do tempo de observação, possibilitando a identificação de diferenças tanto precoces quanto tardias nas plantas.

Modelos que incorporam a influência das covariáveis sobre o tempo até a ocorrência do evento de interesse devem ser analisados com cautela. Uma problemática que surge e que pode influenciar nos modelos é a natureza bienal da produção de café. Essa característica tende a contribuir para que a presença de um grande número de plantas atinja o evento de interesse em um mesmo período, resultando em muitos empates nos tempos observados.

Para lidar com essa questão, serão testados modelos que utilizam aproximações da função de verossimilhança parcial para diferentes situações de empates. As aproximações propostas por Peto (1972), Breslow (1974), Efron (1977) assim como modelos para dados agrupados, buscando maior precisão nas estimativas, serão utilizados nesta etapa.

Deste modo, este trabalho tem como principais objetivos: i) utilizar a análise de sobrevivência para gerar informações sobre os dados fenotípicos de uma população de uma nova espécie de *Coffea arabica*; ii) aplicar os testes de log-rank e Wilcoxon para comparar as curvas de produtividade de diferentes grupos das covariáveis em estudo; iii) ajustar um modelo de riscos proporcionais de Cox, incluindo as covariáveis significativas identificadas; iv) comparar diferentes métodos de aproximação para função de verossimilhança para lidar com empates no modelo de Cox, além de explorar o uso de modelos lineares generalizados que considerem dados agrupados em cenários com múltiplos empates; v) confrontar os resultados obtidos com a análise fenotípica previamente realizada por Mariz et al. (2025), aprofundando o entendimento sobre os padrões observados.

## 2 REFERENCIAL TEÓRICO

### 2.1 Análise de sobrevivência

A análise de sobrevivência surgiu com o intuito de estudar a mortalidade relacionada a pesquisas biomédicas. O primeiro estudo a utilizar tal formato foi publicado por John Graunt em 1662. O trabalho pioneiro foi intitulado *Natural and Political Observations Made Upon the Bills of Mortality*, em que o autor calculou a taxa de mortalidade dos indivíduos a partir de relatórios semanais que detalhavam os números e as causas das mortes no subúrbio de Londres (GUO, 2010).

Segundo Colosimo e Giolo (2006) a análise de sobrevivência ou confiabilidade é a área da estatística na qual o interesse é estimar o tempo até um determinado evento, cuja principal característica, é a presença de dados censurados.

É importante destacar alguns termos comumente utilizados na análise de sobrevivência. Por tempo, queremos dizer anos, meses, semanas ou dias desde o início do acompanhamento de um indivíduo até que um evento ocorra (KLEINBAUM e KLEIN, 1996).

Por eventos, entende-se ocorrências como: morte; incidência de doença; recaída de remissão; falha de uma máquina; tempo até a germinação de sementes; tempo para a colheita de uma safra; ou qualquer outra experiência de interesse que possa ocorrer a um indivíduo, conforme indicam Collett (1993) e Kleinbaum e Klein (1996). Diversos exemplos adicionais podem ser encontrados em Collett (1993), que aborda diferentes tipos de eventos em variadas áreas de estudo.

De acordo com Colosimo e Giolo (2006), a falha é um termo comumente utilizado para indicar que o indivíduo apresentou o evento de interesse. Tal nomenclatura está associada ao fato de que, historicamente, a maioria dos eventos consistia em desfechos indesejados, como morte, incidência de doenças ou outras experiências negativas.

No entanto, é importante destacar que nem sempre a falha será associada a algo negativo. Kleinbaum e Klein (1996) exemplificam que o tempo de sobrevivência pode representar, por exemplo, o tempo para retornar ao trabalho após um procedimento cirúrgico eletivo; nesse caso, a falha corresponde a um evento positivo.

Neste trabalho, entende-se por falha as plantas que atingiram uma produtividade igual ou superior à média anual das testemunhas. A censura corresponde às plantas que morreram durante o período de acompanhamento ou que não apresentaram a falha até o final desse período.

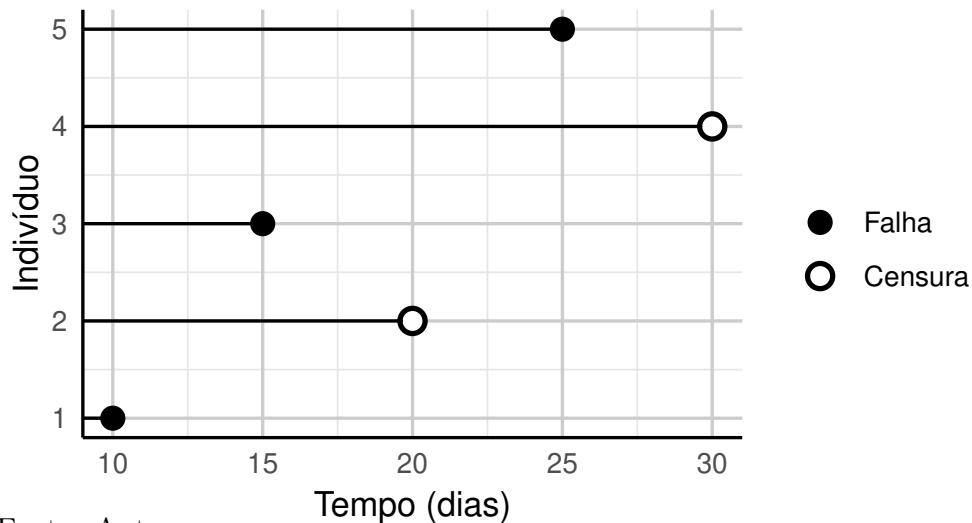
### 2.2 Censura

Os estudos em análise de sobrevivência em geral, envolvem o acompanhamento ao longo do tempo, com o intuito de observar uma resposta durante o período pesquisado. No entanto, o estudo pode terminar antes que ocorra o evento de interesse para todos os indivíduos da amostra. Essa presença de observações incompletas ou parciais ao longo do tempo é chamada de censura. Colosimo e Giolo (2006) definem a censura como a observação parcial da resposta e destacam dois pontos importantes que justificam sua utilização: (i) mesmo incompletas, as observações fornecem informações importantes desde o início do estudo até a censura das observações; (ii) ao omitirmos a presença de censura, podemos obter estatísticas viesadas.

Há três formas distintas de censura: a censura à direita, a censura à esquerda e a censura do tipo intervalar. No caso da censura à direita, ela se desdobra em três cenários distintos, os quais são resumidos por Colosimo e Giolo (2006), da seguinte forma:

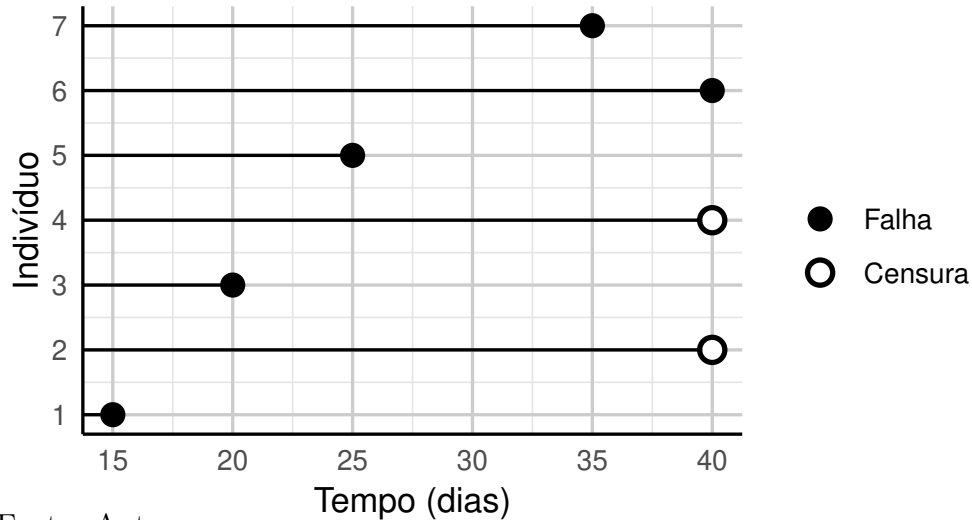
- i) Censura tipo I: Ocorre quando o estudo é encerrado após um período preestabelecido de tempo. Para ilustrar a situação, a Figura 1 representa 5 indivíduos em um estudo fictício com término em 30 dias. Os símbolos • (bola fechada) apresenta o indicativo de falha e o (bola aberta) o de censura serão utilizados nas Figuras de 1 a 3 a seguir. Entende-se por “falha”, como a observação do evento de interesse.

Figura 1 – Mecanismo de censura tipo I.



- ii) Censura tipo II: O estudo tem um número preestabelecido de observações em que se tem interesse de observar a ocorrência do evento sob estudo. A ilustração da Figura 2 apresenta uma situação fictícia de um estudo com 7 indivíduos, no qual o término ocorre quando a quinta falha é registrada.

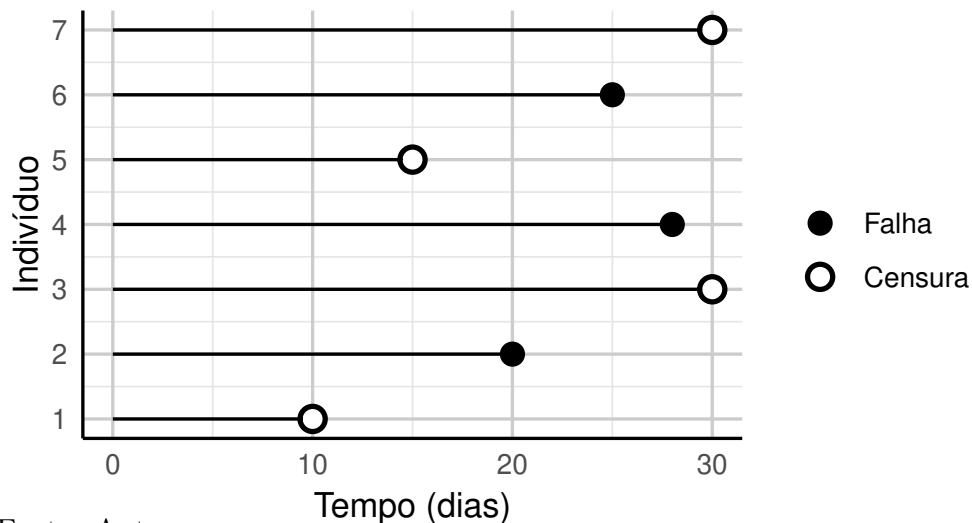
Figura 2 – Mecanismo de censura tipo II.



Fonte: Autor.

iii) Censura tipo III: Acontece quando a observação é retirada do estudo sem que tenha ocorrido a falha. Por exemplo se o paciente é retirado do estudo por uma causa diferente da estudada, como a morte por algum motivo não associado a análise. Como ilustração, a Figura 3 apresenta um estudo fictício com 7 observações, e duração de 30 dias.

Figura 3 – Mecanismo de censura tipo III.



Fonte: Autor.

A censura à esquerda, ocorre quando o tempo registrado é maior do que o tempo de falha, se assemelhando com a censura à direita. Turkson; Ayiah-Mensah e Nimoh (2021) apresentam uma definição mais abrangente dos tipos de censura, em que abordam a relação entre as duas.

**Definição 1.** As variáveis de sobrevivência  $Y_1, \dots, Y_n$  são censuradas à direita pelas constantes fixas,  $t_1, \dots, t_n$  se a amostra observada consistir nos pares ordenados  $(Z_i, \delta_i)$  para  $i = 1, \dots, n$ . Para cada  $i : Z_i = \min \{Y_i, t_i\}$ , no qual  $\delta_i = 1$  se  $Y_i \leq t_i$  caso sem censura, ou

$\delta_i = 0$  se  $Y_i > t_i$  quando o indivíduo foi censurado, em que  $t_i$  é o tempo de censura fixo e  $\delta_i$  é o indicador de censura para o tempo  $T_i$ . Para a censura à esquerda, os tempos observados são  $Z_i = \max\{Y_i, l_i\}$  em que  $l_i$  é o tempo de censura à esquerda associado a  $Y_i$ . Para dados censurados à esquerda,  $-Z_i = \min\{-Y_i, -l_i\}$ .

Um tipo mais geral de censura ocorre quando o tempo de sobrevivência de um indivíduo  $T_i$  ocorre entre dois valores, isto é  $T_i \in [L_i, U_i]$ , sendo que  $L_i \leq T_i \leq U_i$ . Para indivíduos em que o evento de interesse ocorreu entre esses dois valores, sabe-se que o seu respectivo tempo de sobrevivência é no máximo,  $U_i$  e, no mínimo  $L_i$ . Por outro lado, quando se analisa indivíduos cujo evento de interesse não ocorreu até o último momento de coleta, o tempo de sobrevivência pertence ao intervalo  $[L_i, +\infty)$  em que  $L_i$  é o tempo desde a entrada no estudo até o último contato mantido com o mesmo e  $U_i \rightarrow +\infty$  (COLOSIMO e GIOLO, 2006).

Pode-se perceber que, tanto tempos censurados à esquerda como à direita, são casos particulares da censura dita intervalar. Este tipo de censura é muito comum quando os estudos possuem limitações na coleta de dados, sendo um exemplo experimentos agrônômicos, devido a longa periodicidade de coleta dos dados (COLOSIMO; CHALITA e DEMÉTRIO, 2000).

## 2.3 Funções do tempo de sobrevivência

Considere a variável aleatória  $T$ , que assume valores não negativos, sendo comumente contínua e representando o tempo de falha. Geralmente, sua descrição é fornecida por meio da função de sobrevivência ou pela taxa de falha. As Seções 2.3.1 e 2.3.2 apresentarão estas funções, bem como outras funções relacionadas, que são amplamente utilizadas na análise de sobrevivência.

### 2.3.1 Função de sobrevivência

A função de sobrevivência é utilizada para determinar a probabilidade de uma observação não falhar até um certo tempo  $t$ , ou seja, a probabilidade de uma observação sobreviver por mais que um tempo  $t$ . Colosimo e Giolo (2006) definem a função de sobrevivência da seguinte forma:

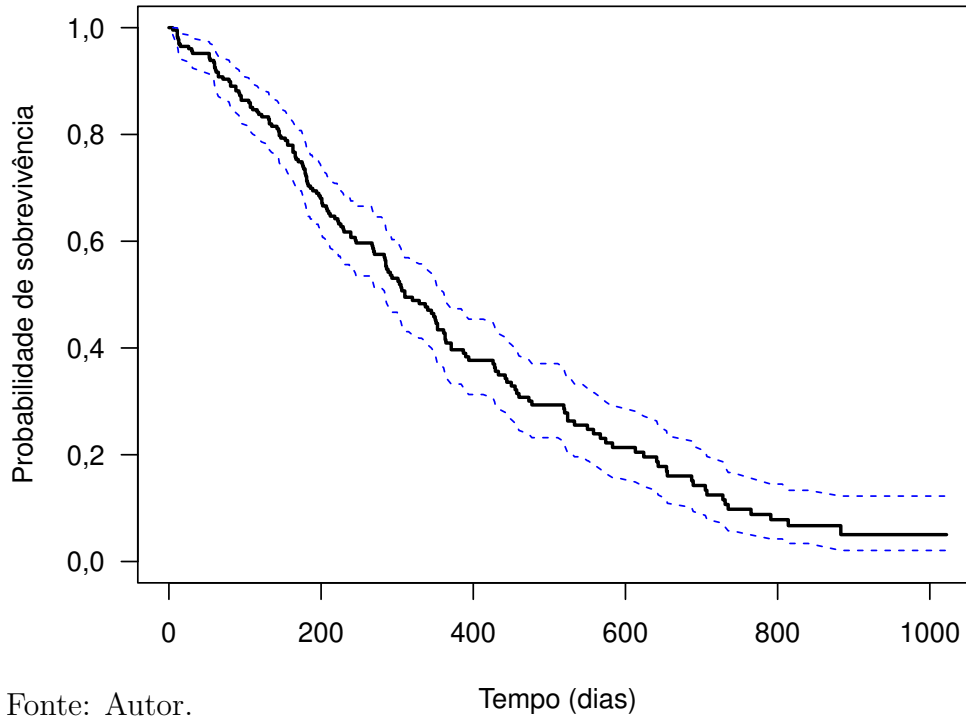
**Definição 2.** Suponha que a variável aleatória  $T$  tenha densidade de probabilidade  $f(t)$ , e função de distribuição,  $F(t)$ . A função de sobrevivência, denotada por  $S(t)$  é definida como a probabilidade da sobrevivência ser maior ou igual do que um certo tempo  $t$ , denotada por,  $S(t) = P(T \geq t)$ . Em consequência a função de distribuição acumulada é definida como a probabilidade de uma observação não sobreviver até o tempo  $t$ , isto é,  $F(t) = 1 - S(t)$ .

A função de sobrevivência é definida para valores não negativos de  $t$ , caracterizando-se por ser não crescente ao longo do tempo. Além disso, a probabilidade de uma observação sobreviver quando o tempo  $t$  tende a zero é igual a um. Segundo Lawless (2003) a função de sobrevivência tem as seguintes propriedades;

- O intervalo de definição do tempo é  $t \in [0, +\infty)$ ;
- $S(t)$  é não crescente, ou seja,  $S(t_1) \geq S(t_2)$ , para  $t_1 \leq t_2$ ;
- $S(t) = 0$  quando  $t \rightarrow +\infty$ .

Na Figura 4 temos um gráfico típico de uma função de sobrevivência estimada a partir do estimador de Kaplan-Meier, o qual será visto na Seção 2.3.4.

Figura 4 – Sobrevivência e respectivos intervalos de 95% de confiança.



Fonte: Autor.

### 2.3.2 Função de risco

Uma função bastante utilizada na análise de sobrevivência devido a sua facilidade de interpretação é a função de risco, comumente denotada por  $\lambda(t)$ . Também conhecida como taxa de falha, ela representa a probabilidade de que a falha ocorra no intervalo  $[t_1, t_2]$ , dado que não ocorreu antes de  $t_1$ , dividida pelo comprimento do intervalo. Ao tomarmos um intervalo com variação pequena, sendo ela  $\Delta t$ , temos a taxa de falha instantânea definida por Colosimo e Giolo (2006) da seguinte forma:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

As principais propriedades da função de risco são:  $\lambda(t) \geq 0$  para todo  $t$ ;  $\lambda(t)$  não tem limite superior e pode assumir qualquer forma. A função de risco pode ainda ser encontrada utilizando a função de distribuição ou a função densidade de probabilidade  $f(t)$ . A relação entre elas origina-se da definição da função de risco para uma taxa instantânea, em seguida aplica-se as propriedades de probabilidade condicional, e por fim utiliza-se a definição de função de sobrevivência (EMMERT-STREIB e DEHMER, 2019).

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{S(t)\Delta t} \\ &= \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}. \end{aligned}$$

### 2.3.3 Função taxa de falha acumulada

A função taxa de falha acumulada, comumente denotada por  $\Lambda(t)$  é útil na avaliação da função de maior interesse que é a função de taxa de falha,  $\lambda(t)$  como indicam os autores COLOSIMO e GIOLO (2006) ainda de acordo com os autores isso acontece essencialmente na estimação não paramétrica, no qual  $\Lambda(t)$  apresenta um estimador com propriedades ótimas e  $\lambda(t)$  é de difícil estimação. De acordo com Collett (1993) a função taxa de falha acumulada no tempo  $t$  pode ser interpretada como o número esperado de eventos que ocorrem no intervalo da origem do tempo até  $t$ . Segundo Colosimo e Giolo (2006) a função é dada por;

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

Outra forma de se chegar na função taxa de falha acumulada é através da relação apresentada por Emmert-Streib e Dehmer (2019);

$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (1)$$

Sabemos que,  $S(t) = 1 - F(t)$ , derivando em ambos os lados, encontramos  $f(t) = -\frac{d}{dt}S(t)$ , daí, substituindo na Equação (1), temos;

$$\lambda(t) = \frac{-\frac{d}{dt}S(t)}{S(t)} = -\frac{d}{dt} \log S(t),$$

agora, integrando ambos os lados em relação a  $t$  e multiplicando por  $-1$  em ambos os lados, encontramos a relação desejada;

$$\Lambda(t) = -\log S(t). \quad (2)$$

Outra relação interessante ocorre ao aplicar a função exponencial em ambos os lados da Equação (2), obtendo-se;

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u) du\right). \quad (3)$$

Outra relação importante referente a uma estimativa da função de taxa de falha acumulada é proposta por Breslow (1972), sendo a estimativa uma função escada com saltos nos tempos distintos de falhas e expressa por:

$$\hat{\lambda}_0(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{l \in R_j} \exp\{\mathbf{x}_l' \boldsymbol{\beta}\}}, \quad (4)$$

em que  $d_j$  é o número de falhas em  $t_j$ ,  $\mathbf{x}_l$  representa o vetor de variáveis no modelo e  $\boldsymbol{\beta}$  representa os coeficientes de regressão.

### 2.3.4 Estimador de Kaplan-Meier (K-M)

Proposto por Kaplan e Meier (1958) esse estimador se destaca pela facilidade de estimação e pela ampla utilização na análise de sobrevivência. É uma técnica não paramétrica, cujo objetivo é descrever o comportamento da função de sobrevivência para diferentes grupos de indivíduos. É uma técnica empírica que se baseia na ocorrência dos eventos e nas censuras ao longo do tempo. D'Arrigo *et al.* (2021) definem o estimador conforme a Definição 3.

**Definição 3.** Suponha que  $d_j$  indique o número de indivíduos que apresentam o evento de interesse no tempo  $t_j$  e  $n_j$  indique o número de indivíduos sob risco no tempo  $t_j$ . A probabilidade  $\hat{q}$  condicional do indivíduo morrer no tempo  $t_j$  dado que ele esteja vivo no tempo  $t_{j-1}$  é;

$$\hat{q} = \frac{d_j}{n_j}.$$

A probabilidade de que o indivíduo sobreviva no tempo  $t_j$  é dada por:

$$\hat{p}_j = 1 - \hat{q}_j = \frac{n_j - d_j}{n_j}.$$

Multiplicando as estimativas das probabilidades condicionais de sobrevivência, obtemos a estimativa da probabilidade cumulativa de sobreviver além do instante  $t_j$ , dada por:

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \hat{p}_j.$$

### 2.3.5 Tabela de vida

A tabela de vida é um dos instrumentos estatísticos mais antigos, e sua construção consiste em dividir o eixo do tempo em um certo número de intervalos. Para cada intervalo estima-se o valor da taxa de falha e a partir da obtenção destes valores, estima-se a função de sobrevivência. De acordo com Colosimo e Giolo (2006) podemos definir a tabela de vida da seguinte forma:

**Definição 4.** Suponha que o eixo do tempo seja dividido em  $s$  intervalos definidos pelos pontos de corte,  $t_1, \dots, t_s$ . Isto é  $I_j = [t_{j-1}, t_j)$ , para  $j = 1, \dots, s$ , em que  $t_0 = 0$  e  $t_s = +\infty$ . A taxa de falha  $\hat{q}_j$  é dada por:

$$\hat{q}_j = \frac{\text{N}^\circ \text{ de falhas no intervalo } [t_{j-1}, t_j)}{[\text{n}^\circ \text{ sob risco em } t_{j-1}] - \left[\frac{1}{2} \times \text{n}^\circ \text{ de censura em } [t_{j-1}, t_j)\right]}.$$

O termo que aparece dividido por dois no denominador é explicado pelo fato de que observações censuradas no intervalo  $[t_{j-1}, t_j)$  são tratadas como se estivessem sob risco durante metade do intervalo considerado. O estimador de sobrevivência da tabela de vida fica então expresso por:

$$\widehat{S}(t) = \prod_{i=1}^j (1 - \hat{q}_{i-1}), \quad t \in I_j,$$

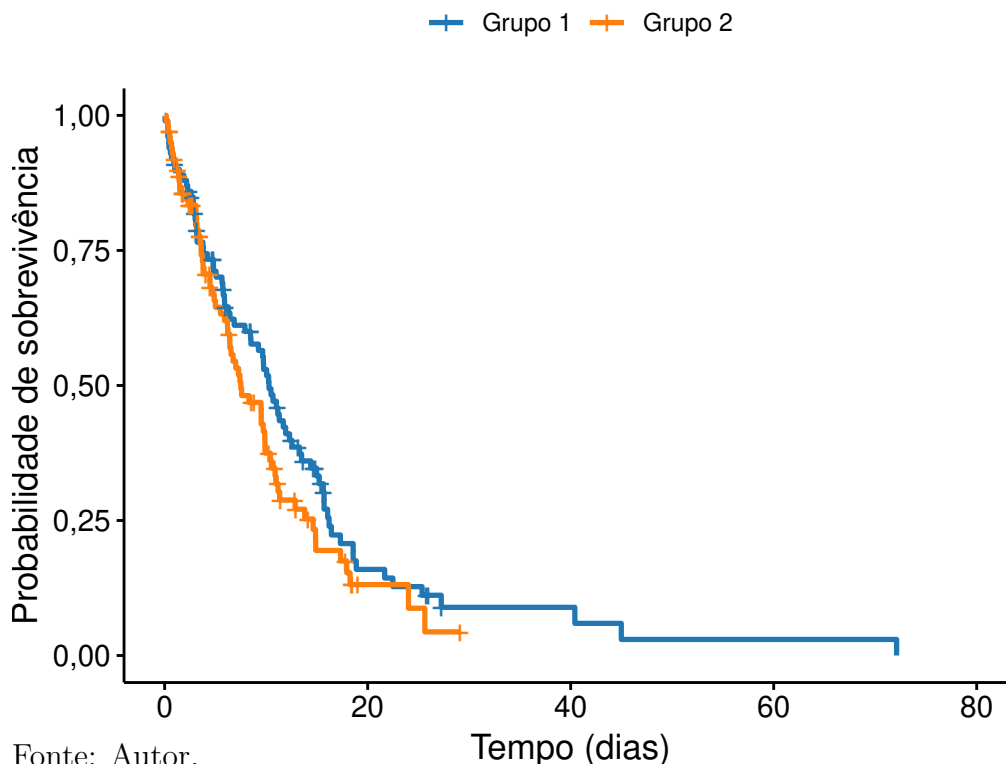
para  $j = 1, \dots, s$  e  $\hat{q}_0 = 0$ .

## 2.4 Comparação de curvas de sobrevivência

Em situações em que existem múltiplas curvas de sobrevivência, é comum surgir o interesse em realizar comparações entre elas. Para ilustrar, considere que temos duas curvas

de sobrevivência que representam dois conjuntos distintos de indivíduos. Por exemplo, em um estudo clínico, suponhamos que existam dois tipos de tratamentos para combater uma doença. A análise comparativa dessas curvas pode fornecer *insights* valiosos sobre a eficácia do melhor tratamento. Para ilustrar, a Figura 5 representa um estudo fictício de 80 dias, comparado dois grupos, um que recebe e outro que não recebe um remédio para uma doença.

Figura 5 – Curvas de sobrevivência: Grupo 1 versus Grupo 2.



Fonte: Autor.

Estatisticamente, uma comparação pode ser realizada por um teste de hipóteses, testando a hipótese nula  $H_0$  : de que não existe diferença na sobrevivência dos grupos *versus* a hipótese alternativa  $H_1$  : de que existe uma diferença significativa na sobrevivência dos grupos (EMMERT-STREIB e DEHMER, 2019).

De acordo com Mantel (1966) o teste de escores é o mais utilizados na literatura. Destaca-se o teste de Wilcoxon, um caso particular do teste de escores, o qual é também, bastante utilizado. A diferença entre os dois testes é que um teste de escores é mais poderoso do que um teste de Wilcoxon para detectar diferenças tardias nas curvas de sobrevivência, enquanto o teste de Wilcoxon é mais poderoso que o teste de escores para detectar diferenças precoces (EMMERT-STREIB e DEHMER, 2019).

### 2.4.1 Teste log-rank

O teste de escores, também conhecido como teste de log-rank de Mantel-Haenszel, em homenagem aos autores, é um teste de hipóteses não paramétrico utilizado para comparar as distribuições de sobrevivência de duas ou mais amostras. Emmert-Streib e Dehmer (2019) indicam as seguintes pressuposições para se utilizar o teste:

- i) Indivíduos censurados e não censurados têm a mesma probabilidade de experimentarem o evento (a censura não é informativa);
- ii) As curvas de Kaplan-Meier dos dois grupos não devem se cruzar (a suposição de riscos proporcionais deve ser mantida).
- iii) Nenhuma distribuição particular para a curva de sobrevivência é assumida (livre de distribuição).

Conforme fornecido por Collett (1993) e Colosimo e Giolo (2006) o teste é definido como sendo:

**Definição 5.** Considere, inicialmente o teste de igualdade de duas funções de sobrevivência  $S_1(t)$  e  $S_2(t)$ , ou seja,  $H_0 : S_1(t) = S_2(t)$ , para todo  $t \in (0, T^*)$ , com  $T^*$  o maior tempo de acompanhamento. Sejam  $t_1 < \dots < t_k$  os tempos distintos e ordenados de falha da amostra formada pela combinação das duas amostras individuais. Para a amostra combinada, suponha que no tempo  $t_j$  ocorram  $d_j$  falhas e que  $n_j$  indivíduos estejam sob risco no tempo imediatamente anterior a  $t_j$ ,  $j = 1, \dots, k$ .

Para a amostra  $i, i = 1, 2$ , suponha, ainda, que ocorram  $d_{ij}$  falhas em  $t_j$  e que  $n_{ij}$  indivíduos estejam sob risco no tempo imediatamente anterior a  $t_j$ . Assim, para cada tempo de falha  $t_j, j = 1, \dots, k$ , pode se dispor os dados em uma tabela de contingência  $2 \times 2$  com  $d_{ij}$  falhas e  $n_{ij} - d_{ij}$ , sobreviventes na coluna  $i, i = 1, 2$ , como mostrado na Tabela 1.

Tabela 1 – Tabela de contingência associada ao tempo de falha  $t_j$

Evento	Grupo 1	Grupo 2	Totais
Falha	$d_{1j}$	$d_{2j}$	$d_j$
Não falha	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
Totais	$n_{1j}$	$n_{2j}$	$n_j$

Condicional à experiência de falhas e censuras até o tempo  $t_j$  (marginais coluna fixa) e ao número de falhas no tempo  $t_j$  (marginais linha fixa), a distribuição de  $d_{2j}$  é a hipergeométrica que é definida por (Mood; GRAYBILL, BOES 1974). Considere  $D_{2j}$  como sendo a variável aleatória em questão que assume valores no intervalo  $[0, \min\{d_j, n_{2j}\}]$ . Nessas condições,  $D_{2j}$  segue uma distribuição hipergeométrica, dada por:

$$f_{D_{2j}(d_{2j}, n_j, d_j, n_{2j})} = \frac{\binom{d_j}{d_{2j}} \binom{n_j - d_j}{n_{2j} - d_{2j}}}{\binom{n_j}{n_{2j}}}$$

e,

$$E(D_{2j}) = e_{2j} = \frac{n_{2j} d_j}{n_j},$$

$$V(D_{2j}) = (V_j)_2 = v_{2j} = n_{2j}(n_j - n_{2j})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1},$$

em que a média de  $e_{2j}$  e  $(V_j)_2$  são a média e a variância da variável  $D_{2j}$ . Então a estatística  $d_{2j} - e_{2j}$  tem média zero e variância  $(V_j)_2$  se as  $k$  tabelas de contingência forem independentes,

um teste aproximado para a igualdade das duas funções de sobrevivência pode ser baseado na estatística:

$$T_L = \frac{\left[\sum_{j=1}^k (d_{2j} - e_{2j})\right]^2}{\sum_{j=1}^k ((V_j)_2)}, \quad (5)$$

que sob  $H_0$ , para todo  $t$  no período de acompanhamento, tem distribuição qui-quadrado com 1 grau de liberdade para grandes amostras.

De acordo com Collet (1993) é possível mostrar que  $\left[\sum_{j=1}^k (d_{2j} - e_{2j})\right] = U_L$  converge assintoticamente para a distribuição normal quando o número de tempos de falha não é baixo. Sob a hipótese de igualdade das curvas de sobrevivência, a razão  $\frac{U_L}{\sqrt{((V_j)_2)}}$  converge para distribuição normal padrão. Sabe-se que se  $Z \sim N(0, 1)$ , então  $Z^2 \sim \chi_1^2$ . Deste modo:

$$W_L = \frac{(U_L)^2}{((V_j)_2)} \sim \chi_1^2.$$

Em situações onde se comparam mais de duas curvas de sobrevivência pode-se utilizar o teste do qui-quadrado, que considera somente os valores observados e esperados das falhas para cada grupo, sem o cálculo da variância da equação (HOSMER; LEMESHOW, 1999).

#### 2.4.2 Teste de Wilcoxon

Proposto por Gehan (1965) o teste de Wilcoxon, segue como um caso particular do teste de log-rank. Ao se considerar o caso particular da comparação de duas funções de sobrevivência,  $S_1(t)$  e  $S_2(t)$  a forma geral a seguir inclui os testes mais importantes na literatura e generaliza a estatística  $T_L$  na Equação 5, (COLOSIMO e GIOLO, 2006).

$$S = \frac{\left[\sum_{j=1}^k u_j (d_{2j} - w_{2j})\right]^2}{\sum_{j=1}^k u_j^2 (V_j)_2}, \quad (6)$$

em que  $u_j$  representa os pesos que especificam os testes.

Sob a hipótese nula de igualdade das curvas de sobrevivência, a estatística  $S$  segue uma distribuição qui-quadrado com 1 grau de liberdade para grandes amostras. O teste de Wilcoxon é obtido ao se tomar  $u_j = n_j$  na Equação 6, enquanto que, ao se adotar  $u_j = 1$ , obtém-se o teste de log-rank. Mais informações sobre esses e outros testes podem ser encontradas em Colosimo e Giolo (2006), bem como em Hosmer e Lemeshow (1999).

## 2.5 Modelo de riscos proporcionais de Cox

Até o momento, foram considerados apenas modelos que não incluíam quaisquer covariáveis dos indivíduos. No entanto, estudos de análise de sobrevivência frequentemente envolvem covariáveis que podem estar relacionadas ao tempo de sobrevivência. Quando se incluem covariáveis na análise, o modelo resultante é denominado Modelo de Riscos Proporcionais de Cox (MRPC). Ele recebe esse nome devido à propriedade de que diferentes indivíduos possuem funções de risco proporcionais, de modo que, ao se tomar a razão dessas funções de

risco, ela permanece constante ao longo do tempo. Caso essa pressuposição não seja atendida, outros modelos devem ser utilizados.

Segundo Cox e Snell (1968) o MRPC é um modelo de regressão semiparamétrico que define a função de risco. Assume-se que os tempos  $t_i$ ,  $i = 1, \dots, n$  são independentes e que a função de taxa de falha do indivíduo  $i$ , dado o vetor  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  de covariáveis, é dada por:

$$h(t|\mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i) \quad (7)$$

em que,  $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$  é o vetor de parâmetros desconhecidos,  $h_0(t)$  é o componente não paramétrico conhecido como função de base, pois  $h(t) = h_0(t)$  quando  $x = 0$  e  $\mathbf{x}_i$  é o conjunto de covariáveis no modelo para o indivíduo  $i$ .

## 2.6 Verossimilhança parcial

No modelo de regressão de Cox, é necessário analisar os efeitos que os coeficientes exercem sobre a função de risco. Dessa forma, torna-se necessário um método de estimação para realizar inferências a partir do modelo. Entre os métodos de estimação, destaca-se o método da máxima verossimilhança, que tem como ideia central, encontrar os valores dos parâmetros que maximizam a função de verossimilhança dos dados, ou seja, que tornam os dados observados mais prováveis dentro do modelo considerado. Entretanto, a presença de um componente não paramétrico na função  $h_0(t)$  torna esse método inadequado. Diante desse problema, Cox (1975) propôs o método de verossimilhança parcial, que condiciona a verossimilhança à história dos tempos de sobrevivência e censuras anteriores, eliminando assim a função de base desconhecida.

A função de verossimilhança parcial  $L(\boldsymbol{\beta})$  proposta por Cox (1975), leva em conta diversas ponderações, sendo dada pela Definição 6:

**Definição 6.** Considere que em uma amostra com  $n$  indivíduos, existam  $k \leq n$  falhas distintas nos tempos  $t_1 < \dots < t_k$ . A probabilidade condicional da  $i$ -ésima observação vir a falhar no tempo  $t_i$  conhecendo quais observações estão sob risco em  $t_i$  é:

$$\frac{h_i(t|\mathbf{x}_i)}{\sum_{j \in R(t_i)} h_j(t|\mathbf{x}_j)} = \frac{h_0(t) \exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} h_0(t) \exp(\mathbf{x}'_j \boldsymbol{\beta})} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}'_j \boldsymbol{\beta})}, \quad (8)$$

em que  $R(t_i)$  é o conjunto dos índices das observações sob risco no tempo  $t_i$ . Ao utilizar probabilidade condicional pode-se verificar que o componente não paramétrico  $h_0(t)$  desaparece da Equação (8). A função de verossimilhança parcial denotada por  $L(\boldsymbol{\beta})$  é formada pelo produto das probabilidades condicionais, associadas aos distintos tempos de falhas, isto é,

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}'_j \boldsymbol{\beta})} = \prod_{i=1}^n \left( \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}'_j \boldsymbol{\beta})} \right)^{\delta_i}, \quad (9)$$

na qual  $\delta_i$  é o indicador de falha admitindo valor zero se o  $i$ -ésimo tempo de sobrevivência é censurado e valor um caso contrário.

Os valores de  $\beta$  que maximizam a função  $L(\beta)$ , são obtidos resolvendo-se o sistema de equações, definida por  $U(\beta) = 0$ , em que  $U(\beta)$  é o vetor escore de derivadas de primeira ordem da função suporte, isto é,  $\ell(\beta) = \ln(L(\beta))$ , assim,

$$U(\hat{\beta}) = \sum_{i=1}^n \delta_i \left[ x_i - \frac{\sum_{j \in R(t_i)} x_j \exp \{x'_j \hat{\beta}\}}{\sum_{j \in R(t_i)} \exp \{x'_j \hat{\beta}\}} \right] = 0.$$

Na Equação (9) a função de verossimilhança parcial assume que os tempos de sobrevivência são contínuos e, conseqüentemente não pressupõe a possibilidade de empates nos valores observados (COLOSIMO e GIOLO, 2006). Entretanto, devido a escala de medida utilizada, em situações práticas os empates podem acontecer nos tempos de falhas ou censuras, da mesma forma que podem acontecer empates entre falhas e censuras. Quando ocorrem empates entre falhas e censuras, ou seja, quando os tempos de falhas são iguais, mas um deles é censurado, por convenção, considera-se que a censura ocorreu após a falha.

## 2.7 Aproximação da verossimilhança parcial

Na literatura, as três aproximações mais comumente utilizadas para a função de verossimilhança parcial são as propostas por Breslow (1974), Efron (1977) e a aproximação na forma exata, proposta por Peto (1972). De acordo com Collet (1993) e Colosimo e Giolo (2006), a aproximação exata se caracteriza por um melhor ajuste em comparação às demais. No entanto, ela é computacionalmente mais robusta, o que torna os cálculos estatísticos extremamente complexos quando o número de empates é grande em qualquer tempo.

Chalita (1997) destaca que a aproximação de Efron é mais próxima da exata do que a de Breslow. Além disso, Chalita *et al.* (2002) observam que, em média, a aproximação de Breslow tende a subestimar os parâmetros em relação à aproximação de Efron. Picciotto e Rockhill (1997) e Chalita *et al.* (2002) recomendam o uso da aproximação de Efron em vez da aproximação de Breslow, uma vez que a primeira é ligeiramente mais robusta e apresenta um melhor ajuste.

De acordo com Colosimo e Giolo (2006), pela simplicidade de implementação e por ser facilmente encontrada nos *softwares* estatísticos, a aproximação de Breslow é a mais utilizada.

Atualmente, as três aproximações estão disponíveis em diversos *softwares* estatísticos como o R (2025) que ajustam o modelo de taxa de falhas proporcionais, permitindo ao pesquisador escolher a mais adequada para o seu contexto de análise.

### 2.7.1 Aproximação exata

Originalmente apresentada por Peto (1972), Colosimo e Giolo (2024) apresentam a definição original da seguinte forma:

**Definição 7.** Considere  $d_j$  falhas no tempo  $t_j$ ,  $j = 1, \dots, k$ , em que  $t_1 < t_2 < \dots < t_k$ . Embora o posto dos tempos de falha em  $t_j$  seja menor do que o daqueles em  $t_{j+1}$ , o arranjo da ordenação das  $d_j$  falhas é desconhecido. Considere cada uma das possibilidades por  $A_s$ , com

$s = 1, \dots, d_j!$  e  $G_j = P(A_1 \cup \dots \cup A_{d_j!}) = \sum_{s=1}^{d_j!} P(A_s)$ . Desse modo, a função de verossimilhança parcial exata considerando empates é dada por:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^k G_j = \prod_{j=1}^k \left[ \exp(\boldsymbol{\beta}' \mathbf{s}_j) \sum_{P \in Q_j} \prod_{r=1}^{d_j} \left( \sum_{l \in R(t_j, p_r)} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right)^{-1} \right], \quad (10)$$

em que  $Q_j$  é o conjunto das permutações dos índices  $1, \dots, d_j$ ,  $P = (p_1, \dots, p_{d_j})$  é um elemento de  $Q_j$ ,  $R(t_j, p_r)$  é o conjunto da diferença  $R(t_j) - \{p_1, \dots, p_{r-1}\}$ , e  $\mathbf{s}_j = \sum_{l \in D_j} \mathbf{x}_l$  e  $D_j$  é o conjunto dos indivíduos que falharam em  $t_j$ .

### 2.7.2 Aproximação de Efron

A aproximação da função de verossimilhança parcial exata dada pela Expressão (10) fica um tanto quanto complexa se o número de empates for grande em qualquer tempo de falha (COLOSIMO E GIOLO, 2024). Uma alternativa foi apresentada por Efron (1977), que adiciona a pressuposição de que a razão de riscos é constante em qualquer intervalo de tempo, em que o conjunto de indivíduos sob risco não muda no intervalo, exceto quando ocorre a falha. De acordo com Efron (1977) a aproximação alternativa é então dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta}' \mathbf{s}_i)}{\prod_{r=1}^{d_i} \left[ \left( \sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right) - (r-1) d_i^{-1} \left( \sum_{l \in D_i} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right) \right]},$$

em que  $D_i$  é o conjunto de indivíduos que falharam em  $t_i$ .

### 2.7.3 Aproximação de Breslow

Uma generalização da aproximação proposta por Peto foi realizada por Breslow (1974), onde o autor propõe uma versão discreta do modelo de riscos proporcionais de Cox (1972) que considera a partição de um espaço amostral contínuo (CHALITA, 1997). A aproximação proposta é dada por:

**Definição 8.** Considere  $\mathbf{s}_i$  o vetor formado pela soma das correspondentes  $p$  covariáveis para os indivíduos que falharam no mesmo tempo  $t_i$  ( $i = 1, \dots, k$ ) e  $d_i$  o número de falhas nesse mesmo tempo. A aproximação da função de verossimilhança é expressa por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta}' \mathbf{s}_i)}{\left[ \sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right]^{d_i}}.$$

Quando a presença de empates é pequena, as aproximações de Breslow e Efron resultam em valores muito próximos e, na ausência de empates, ambas se reduzem à função de verossimilhança parcial usual, dada na Expressão (9).

Alguns autores compararam a eficácia dessas aproximações. Por exemplo, Picciotto e Rockhill (1997), por meio de processos de simulação, verificaram que a aproximação de Efron apresenta melhores resultados do que a de Breslow, principalmente quando a quantidade de empates é grande.

Por outro lado, Colosimo e Giolo (2006) destacam que a aproximação de Breslow é mais utilizada, devido à sua simplicidade computacional e à disponibilidade em diversos softwares estatísticos.

## 2.8 Modelos de regressão discretos

Nas situações em que o número de observações empatadas em qualquer tempo não é grande, podem ser encontradas alternativas para a função representada na Expressão (9) em Breslow (1974) e Peto (1972), além de Efron (1977). Quando o número de empates é grande, aconselha-se utilizar o modelo de regressão de Cox para dados agrupados conforme recomendam Lawless (2003), Prentice e Gloeckler (1978). Colosimo; Chalita e Demétrio (2000) com adaptações em Strapsson (2007) abordam essa situação da seguinte maneira:

**Definição 9.** Suponha que as observações são tomadas sobre  $n$  indivíduos, com variável tempo de vida  $T$  e um vetor  $\mathbf{x} = (x_1, \dots, x_p)'$  de covariáveis associadas a cada indivíduo. Os tempos de vida são agrupados em  $k$  intervalos de igual comprimento,  $I_j = [a_{j-1}, a_j]$ ,  $j = 1, \dots, k$ , sendo que  $0 = a_0 \leq a_1 \leq \dots < a_k = +\infty$ . Vamos supor que todas as censuras ocorrem no final dos intervalos. Seja  $R_t$  o conjunto de indivíduos sob risco no tempo  $a_{j-1}$  e  $\delta_{ij}$  é uma variável indicadora em que  $\delta_{ij} = 1$ , se o tempo de vida do indivíduo  $i$  ocorre dentro de  $I_j$  e  $\delta_{ij} = 0$ , caso contrário. Usualmente a função de verossimilhança é escrita em termos da probabilidade de morte do  $i$ -ésimo indivíduo em  $I_j$  condicionado a estar vivo em  $a_{j-1}$  e aos valores das covariáveis, isto é,

$$\begin{aligned} p_j(\mathbf{x}_i) &= P[a_{j-1} \leq T_i < a_j \mid T_i \geq a_{j-1}, \mathbf{x}_i] = \frac{P[\{a_{j-1} \leq T_i < a_j\} \cap \{T_i \geq a_{j-1}\}]}{P[T_i \geq a_{j-1}]} \\ &= \frac{P[a_{j-1} \leq T_i < a_j]}{P[T_i \geq a_{j-1}]} = \frac{S(a_{j-1} \mid \mathbf{x}_i) - S(a_j \mid \mathbf{x}_i)}{S(a_{j-1} \mid \mathbf{x}_i)}, \end{aligned}$$

e portanto,

$$p_j(\mathbf{x}_i) = 1 - \frac{S(a_j \mid \mathbf{x}_i)}{S(a_{j-1} \mid \mathbf{x}_i)}. \quad (11)$$

Deste modo a função de verossimilhança pode ser obtida considerando as covariáveis  $\mathbf{x}_i$  tais que a contribuição da observação não censurada em  $I_j$  para a função de verossimilhança é;

$$\begin{aligned} P[a_{j-1} \leq T_i < a_j] &= S(a_{j-1} \mid \mathbf{x}_i) - S(a_j \mid \mathbf{x}_i) \\ &= \{1 - p_1(\mathbf{x}_i)\} \times \dots \times \{1 - p_{j-1}(\mathbf{x}_i)\} p_j(\mathbf{x}_i), \end{aligned}$$

e a contribuição da observação censurada em  $I_j$  para a função de verossimilhança é;

$$\begin{aligned} P[T_i \geq a_j] &= S(a_j \mid \mathbf{x}_i) \\ &= \{1 - p_1(\mathbf{x}_i)\} \times \dots \times \{1 - p_j(\mathbf{x}_i)\}. \end{aligned}$$

Desta forma, a função de verossimilhança é dada por:

$$\prod_{j=1}^k \prod_{i \in R_j} \{p_j(\mathbf{x}_i)\}^{\delta_{ij}} \{1 - p_j(\mathbf{x}_i)\}^{1 - \delta_{ij}}. \quad (12)$$

A função de verossimilhança, representada pela Equação (12) corresponde a uma distribuição Bernoulli, cuja variável resposta  $\delta_{ij}$  é binária (a falha acontece ou a falha não acontece no intervalo  $I_j$ ) com probabilidade de sucesso  $p_j(\mathbf{x}_i)$  que pode ser modelada considerando modelos de riscos proporcionais de Cox, Weibull, logístico, dentre outros.

### 2.8.1 Modelo de taxas de falhas proporcionais

Na Seção 2.5 foi apresentada a função taxa de falha, representada pela Equação (7), utilizando as Equações (2) e (3) podemos escrever a função de sobrevivência da seguinte forma;

$$\begin{aligned} S(t|\mathbf{x}_i) &= \exp\left(-\int_0^t h(u|x)du\right) = \exp\left[-\int_0^t \exp\{\boldsymbol{\beta}'\mathbf{x}_i\}h_0(u)du\right] \\ &= \exp\left[-\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}\int_0^t h_0(u)du\right] = \left[\exp\left(-\int_0^t h_0(u)du\right)\right]^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}} \\ &= [S_0(t)]^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}. \end{aligned} \quad (13)$$

Na Expressão (13), o termo  $S_0(t)$  representa a função de sobrevivência base. Agora, substituindo-a na Equação (11) temos;

$$p_j(\mathbf{x}_i) = 1 - \left[\frac{S_0(a_j)}{S_0(a_{j-1})}\right]^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}},$$

e a função de verossimilhança é então expressa por;

$$\prod_{j=1}^k \prod_{i \in R_j} \left[\left(1 - \frac{S_0(a_j)}{S_0(a_{j-1})}\right)^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}\right]^{\delta_{ij}} \left[\left(\frac{S_0(a_j)}{S_0(a_{j-1})}\right)^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}\right]^{1 - \delta_{ij}}. \quad (14)$$

Aplicando o logaritmo neperiano na Expressão (14), temos a função suporte, dada por:

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{j=1}^k \sum_{i \in R_j} \left(\delta_{ij} \ln\left(1 - \alpha_j^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}\right) + (1 - \delta_{ij}) \ln\left(\alpha_j^{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}\right)\right), \quad (15)$$

sendo  $\alpha_j = S_0(a_j)/S_0(a_{j-1})$ .

Prentice e Gloeckler (1978) sugeriram o uso da reparametrização  $\gamma_j^* = \log(-\log(\alpha_j))$ , que torna os  $\gamma_j^*$ 's restritos e a convergência do processo iterativo de estimação dos parâmetros mais rápida (COLOSSIMO E GIOLO, 2006).

A Equação (15) reparametrizada é dada por:

$$\ell_C(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{j=1}^k \sum_{i \in R_j} \left[-(1 - \delta_{ij}) \exp(\gamma_i^* + \boldsymbol{\beta}'\mathbf{x}_i) + \delta_{ij} \log\left(1 - \exp\left(-\exp(\gamma_i^* + \boldsymbol{\beta}'\mathbf{x}_i)\right)\right)\right],$$

as equações das derivadas parciais são dadas por:

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\beta}, \gamma_i^*)}{\partial \beta_j} &= \sum_{j=1}^k \sum_{i \in R_j} \left[ \frac{\delta_i x_{ij} \exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i)}{\exp(\exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i)) - 1} - (1 - \delta_i) x_{ij} \exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i) \right] \\ \frac{\partial \ln L(\boldsymbol{\beta}, \gamma_i^*)}{\partial \gamma_i} &= \sum_{i \in R_i} \left[ \frac{\delta_i \exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i)}{\exp(\exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i)) - 1} - (1 - \delta_i) \exp(\gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_i) \right]\end{aligned}$$

### 2.8.2 Modelo logístico

Assumindo um modelo logístico para o tempo de vida  $T$ , tem-se que:

$$p_i(\mathbf{x}_l) = 1 - (1 + \gamma_i \exp\{\mathbf{x}_l \boldsymbol{\beta}\})^{-1}, \quad (16)$$

em que  $\gamma_i = \{1 - p_i(0)\}$ , para  $i = 1, \dots, k$ . O modelo anterior pode ser linearizado utilizando-se uma transformação *logito*, tal que:

$$\log\left(\frac{p_i(\mathbf{x}_l)}{1 - p_i(\mathbf{x}_l)}\right) = \gamma_i^* + \boldsymbol{\beta}' \mathbf{x}_l = \eta_{lj},$$

em que  $\gamma_i^* = \log(\gamma_j)$  é o efeito do  $j$ -ésimo intervalo de tempo.

Agora, substituindo a Equação (16) na Equação (12), obtém-se o modelo de chances proporcionais para dados agrupados. A função suporte para este modelo é dado por:

$$\log[L(\boldsymbol{\theta})] = \sum_{j=1}^k \sum_{l \in R(t_j)} \left[ \log\left([1 + \gamma_j \exp(\boldsymbol{\beta}' \mathbf{x}_l)]^{-1}\right) + \delta_{lj} (\log(\gamma_j) + \boldsymbol{\beta}' \mathbf{x}_l) \right].$$

A reparametrização  $\gamma_j^* = \log(\gamma_j)$  é indicada para acelerar a convergência do método numérico de estimação dos parâmetros, como aponta Lawles (1982), e a função suporte é dada por:

$$\log[L(\boldsymbol{\beta}, \boldsymbol{\gamma}^*)] = \sum_{j=1}^k \sum_{l \in R(t_j)} \left[ -\log[1 + \exp(\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l)] + \delta_{lj} (\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l) \right],$$

e as equações para obtenção dos estimadores de máxima verossimilhança são dadas por:

$$\begin{aligned}\frac{\partial \log[L(\boldsymbol{\beta}, \boldsymbol{\gamma}^*)]}{\partial \beta_q} &= \sum_{j=1}^k \sum_{l \in R(t_j)} \left[ \delta_{lj} x_{lq} - \frac{x_{lq} \exp(\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l)}{1 + \exp(\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l)} \right], \\ \frac{\partial \log[L(\boldsymbol{\beta}, \boldsymbol{\gamma}^*)]}{\partial \gamma_j^*} &= \sum_{j=1}^k \sum_{l \in R(t_j)} \left[ \delta_{lj} - \frac{\exp(\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l)}{1 + \exp(\gamma_j^* + \boldsymbol{\beta}' \mathbf{x}_l)} \right].\end{aligned}$$

## 2.9 Critérios de informação

Durante o processo de seleção de modelos, é fundamental ter em mente que não existem modelos verdadeiros ou mesmo um único modelo ideal que caracterize perfeitamente a variável

resposta em estudo. De acordo com Oyama *et al.* (2020) a modelagem estatística é, em essência, uma aproximação da realidade, e o pesquisador deve buscar o modelo que melhor equilibre ajuste aos dados e simplicidade, evitando tanto o *underfitting* (subajuste) quanto o *overfitting* (sobreajuste).

Nesse contexto, Sobral e Barreto (2011) destacam dois desafios principais enfrentados pelo pesquisador: O primeiro consiste em definir a ordem em que os modelos devem ser considerados, o que envolve a escolha de variáveis, estruturas de dependência e formas funcionais e o segundo é avaliar a adequação do modelo aos dados observados, garantindo que ele capture os padrões relevantes sem se tornar excessivamente complexo.

Uma abordagem comum para superar esses desafios é utilizar critérios de informação, que permitem comparar modelos com base em sua capacidade de explicar os dados, penalizando a complexidade excessiva. Se uma boa estimativa para a log-verossimilhança esperada puder ser obtida a partir dos dados observados, essa estimativa pode ser utilizada como base para a comparação de modelos (EMILIANO, VEIGA e MENEZES; 2010).

Entre as diversas metodologias disponíveis, enfatizaremos, na Seção 2.9.1, os critérios de Akaike (AIC) e o Akaike Corrigido (AICc), que são amplamente utilizados para selecionar modelos com base no equilíbrio entre ajuste e parcimônia. Na Seção 2.9.2, abordaremos o Critério Bayesiano de Informação (BIC), que incorpora uma penalização mais rigorosa para a complexidade do modelo, sendo particularmente útil em amostras grandes.

### 2.9.1 Critério de informação de Akaike - AIC

Proposto por Akaike (1974) o AIC é dado por:

$$AIC = -2\log L(\hat{\theta}) + 2p \quad (17)$$

em que  $L(\hat{\theta})$  é a função de verossimilhança aplicada nos estimadores de máxima verossimilhança para o modelo e  $p$  é o número de parâmetros a serem estimados. A Equação (17) mostra que o AIC recompensa a precisão descritiva por meio da máxima verossimilhança e penaliza a falta de parcimônia de acordo com o número de parâmetros livres. A equação é baseada em aproximações assintóticas e é válida apenas para conjuntos de dados suficientemente grandes. Pensando nisso, Bozdogan (1987) propôs a seguinte correção para o AIC, dando origem ao critério de Akaike conhecido como (*AICc*).

$$AIC_c = -2\log L(\hat{\theta}) + 2p + 2\frac{p(p+1)}{n-p-1}.$$

### 2.9.2 Critério de informação Bayesiano - BIC

O BIC, proposto por Schwarz (1978) é dado por:

$$BIC = -2\log L(\hat{\theta}) + p\log n,$$

em que  $L(\hat{\theta})$  é o valor máximo da função de verossimilhança,  $\hat{\theta}$  são os valores dos parâmetros que maximizam a função de verossimilhança,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra. De acordo com Vriezi (2012), desde que o modelo

correto esteja no conjunto de modelos a ser testado, o BIC é um método consistente, que escolhe o modelo de ordem correta com probabilidade 1, à medida que o tamanho da amostra tende ao infinito.

## 2.10 Metodologia de Collet

Originariamente proposto por Collet (1993), com variações em Hosmer e Lemeshow (1999) e adaptações em Colosimo e Giolo (2006) a estratégia de seleção de variáveis proposta se assemelha ao método *stepwise*, proposto originalmente por Efroymson (1960) entretanto, possibilita uma autonomia maior do pesquisador no processo de seleção. Os passos utilizados por Collett e adaptados por Colosimo e Giolo (2006) para o processo de seleção são apresentados a seguir;

- Passo 1: Consiste em ajustar modelos que contenham cada uma das covariáveis, selecionando para o próximo passo apenas as variáveis que forem significativas ao nível  $\alpha$  de significância desejado pelo pesquisador.
- Passo 2: As covariáveis que foram significativas no passo 1 são ajustadas conjuntamente em um modelo. Nessa nova etapa, são ajustados modelos reduzidos, excluindo-se uma covariável de cada vez. Permanecem no modelo, apenas, aquelas que produzirem um aumento significativo na estatística do teste da razão de verossimilhança.
- Passo 3: Ajusta-se um modelo com as covariáveis que resistiram ao passo 2. Nesta etapa as covariáveis excluídas no passo 2 retornam para o modelo para verificar se realmente não são significativas.
- Passo 4: As covariáveis que foram significativas no passo 3 são mantidas no modelo. Agora, as covariáveis excluídas no passo 1 retornam para o modelo para verificar se realmente não são significativas.
- Passo 5: Ajusta-se um modelo com as covariáveis significativas do passo 4 e em seguida, verifica se alguma delas pode ser excluída do modelo.

## 2.11 Avaliação dos modelos

Antes de interpretar os parâmetros do modelo, é essencial avaliar o comportamento dos resíduos para assegurar sua adequação. Nesta seção, serão apresentados os resíduos de Cox-Snell para verificar a qualidade do ajuste, bem como os resíduos *martingal* que será utilizado para gerar os resíduos de Cox-Snell e de Schoenfeld, que servirão para avaliar a acurácia do modelo e testar a pressuposição de riscos proporcionais. O desempenho dos modelos de regressão binária ponderada foram analisados por meio da curva ROC (Receiver Operating Characteristic), sendo a área sob a curva (AUC) classificada segundo os critérios: inferior a 0,7 como insuficiente, entre 0,7 e 0,8 como aceitável e superior a 0,8 como excelente (HOSMER *et al.* 2013).

### 2.11.1 Resíduos de Cox e Snell

De acordo com Collett (1993) o resíduo mais amplamente utilizado na análise de dados de sobrevivência é o resíduo de Cox-Snell, assim chamado porque é um exemplo particular da definição geral de resíduos dada por Cox e Snell (1968). Colosimo e Giolo (2024) destacam que os resíduos são utilizados para avaliar a qualidade geral do ajuste do modelo de Cox. Os autores, apresentam a definição a seguir:

**Definição 10.** O resíduo de Cox-Snell para o  $i$ -ésimo indivíduo,  $i = 1, 2, \dots, n$ , é dado por;

$$\hat{e}_i = \hat{\Lambda}_0(t_i) \exp \left\{ \sum_{k=1}^p x_{ik} \hat{\beta}_k \right\}, \quad i = 1, \dots, n, \quad (18)$$

com  $\hat{\Lambda}_0(t_i)$  estimado pela Equação (4). Deste modo, se o modelo estiver bem ajustado, os  $\hat{e}_i$ 's devem ser vistos como uma amostra censurada de uma distribuição normal.

Ainda de acordo com Colosimo e Giolo (2024) os gráficos apresentados pelos resíduos não fornecem informações sobre o tipo de problema que estaria impedindo o ajuste de ser satisfatório. Deste modo, para verificar pressuposições de taxas de falhas proporcionais, os resíduos de Cox-Snell não são indicados.

### 2.11.2 Resíduos *martingal*

Os resíduos de *martingal* são definidos por Colosimo e Giolo (2024) pela expressão:

$$\hat{m}_i = \delta_i - \hat{e}_i,$$

em que  $\delta_i$  é a variável indicadora de falha e  $\hat{e}_i$ , os resíduos de Cox-Snell indicados na Expressão (18). Em um contexto geral, Collet (1993) indica que os gráficos desses resíduos em relação ao tempo de sobrevivência, contribuem para indicar se há tempos específicos, ou valores das variáveis explicativas, onde o modelo não se ajusta bem e que merecem algumas considerações ou podem ser consideradas *outliers*.

### 2.11.3 Resíduos Schoenfeld

Uma das propostas mais comuns para se avaliar a suposição de taxas de falhas proporcionais no modelo de Cox, é a de analisar os resíduos de Schoenfeld (1982). A definição que pode ser encontrada em Colett (1993), é dada por;

**Definição 11.** O  $i$ -ésimo resíduo de Schoenfeld para a  $j$ -ésima variável explicativa no modelo denotada por  $X_j$ , é dado por:

$$rs_{ij} = \delta_i \{x_{ji} - \hat{a}_{ji}\}$$

em que  $x_{ji}$  é o valor da  $j$ -ésima variável explicativa,  $j = 1, 2, \dots, p$ , para o  $i$ -ésimo indivíduo no estudo,

$$\hat{a}_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\boldsymbol{\beta}' \mathbf{x}_l)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)},$$

e  $R(t_i)$  é o conjunto de todos os indivíduos em risco no momento  $t_i$ .

Um problema que surge na verificação das suposições de riscos proporcionais ocorre quando variáveis contínuas estão presentes no modelo final. Nesse caso, é essencial verificar se essa pressuposição é atendida. De acordo com Kleinbaum e Klein (1996), uma alternativa consiste em categorizar as variáveis contínuas. Os autores destacam a importância de tomar cuidado para que sejam criadas poucas categorias, assegurando que as escolhas sejam as mais significativas possíveis.

Além dos resíduos Schoenfeld, uma outra proposta existente para verificar as pressuposições de riscos proporcionais, é a estimação da curva do logaritmo da função de risco acumulado versus o tempo, ou versus o logaritmo do tempo, como indicam os autores Kleinbaum e Klein (1996) e Colosimo e Giolo (2024). Ambos os autores reforçam ainda que as curvas devem apresentar diferenças aproximadamente constantes durante o tempo, enquanto que curvas não paralelas significam desvios das suposições.

## 3 MATERIAL E MÉTODOS

### 3.1 Descrição dos dados

Neste trabalho, utilizou-se uma população oriunda do cruzamento das cultivares Tupi Amarelo IAC 5162 e Híbrido de Timor MG0357. A primeira, tem sua origem provavelmente atrelada a um cruzamento natural entre a cultivar Tupi IAC 1669-33 e “Catuaí Amarelo”. Suas principais características são: porte baixo, diâmetro da copa médio, os grãos possuem um tamanho grande e a maturação é precoce. A cultivar ainda apresenta resistência à ferrugem, possui uma alta capacidade de produção e uma qualidade de bebida regular (CARVALHO *et al.*, 2022). Já o Híbrido de Timor MG0357 apresenta boa qualidade de bebida, porte alto, ciclos médios de maturação e alta resistência a doenças e pragas (MARIZ *et al.*, 2025).

De acordo com Mariz *et al.* (2025) o intuito do cruzamento foi piramidar genes de resistência a *Hemileia vastatrix*, fungo causador da ferrugem do café, e resistência a outras doenças e pragas, visando resistência múltipla e duradoura. Outros propósitos do cruzamento foram redução da altura da planta e aproveitamento do potencial de qualidade superior de bebida da parental HdT MG0357, que foi previamente verificado em testes de análise sensorial.

As cultivares utilizadas como testemunhas no experimento foram Paraíso MGH419-1 (apresenta alta resistência a ferrugem, porte baixo, maturação média, alta produtividade e boa qualidade de bebida) e Catuaí Vermelho IAC 144 (produtiva e muito plantada no Brasil) (MARIZ *et al.*, 2025).

Os cruzamentos foram realizados pela equipe técnica do programa de melhoramento genético do cafeeiro conduzido pela Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), em parceria com a Embrapa Café e a Universidade Federal de Viçosa (UFV), em outubro de 2011, no Campo Experimental de Patrocínio (CEPC/EPAMIG), em Patrocínio-MG. Desse cruzamento (Híbrido de Timor MG0357 X Tupi Amarelo IAC 5162) originou-se uma população  $F_1$ , que foi autofecundada, resultando em uma população  $F_2$  composta inicialmente por 200 indivíduos. Esses indivíduos foram distribuídos em quatro blocos, com 50 plantas em cada.

No presente estudo, foram mensuradas oito características relacionadas ao vigor vegetativo, três associadas à sanidade e três relativas à qualidade da bebida. As covariáveis referentes ao delineamento experimental, bem como os dados de temperatura e precipitação, também foram incluídas na análise.

Ressalta-se que, conforme a convenção adotada na área de análise de sobrevivência, utiliza-se o termo covariáveis para se referir às variáveis explicativas que não correspondem à resposta do estudo. Dessa forma, temperatura, precipitação e as características relacionadas ao delineamento experimental serão tratadas como covariáveis. Todas as covariáveis utilizadas estão apresentadas na Tabela 2.

Tabela 2 – Covariáveis utilizadas no estudo.

	Característica	Abreviação
	Fileira	FIL
	Bloco	BLO
	Vigor vegetativo	VIG
	Altura da planta	APL
	Diâmetro da copa	DCO
	Diâmetro do caule	DCA
	Nº de pares de ramos plagiotrópicos na haste principal	NPR
	Comprimento de ramo plagiotrópico representativo da planta	CRP
	Nº de nós no ramo plagiotrópico	NNR
	Incidência de cercospora	CER
	Incidência de ferrugem	FER
	Infestação de bicho-mineiro	BM
	Tamanho dos frutos maduros	TF
	Ciclo de maturação dos frutos	CM
	Uniformidade de maturação dos frutos	UMT
	Produção	PRO
	Temperatura	TEMP
	Precipitação	PREC

As covariáveis apresentadas na Tabela 2 se dividem em dois grupos principais, covariáveis contínuas e covariáveis categóricas. Foram utilizadas no estudo 8 covariáveis categóricas e 10 covariáveis contínuas, como indicado nas Tabelas 3 e 4 respectivamente. Vale a pena destacar que a covariável VIG, é uma covariável categórica, com notas de 1 a 10. Nota 1: Planta totalmente depauperada. Nota 10: planta considerada com vigor máximo. Entretanto, as notas apresentaram uma variação muito alta nos valores, e então consideramos a covariável VIG como contínua.

As covariáveis categóricas são representadas na Tabela 3 a seguir:

Tabela 3 – Covariáveis categóricas.

Abreviação	Níveis
FIL	Fileira onde está localizado o tratamento no ensaio, cujas notas foram atribuídas para cada fileira, da seguinte forma, 1, 2, 3, 4, 5, 6.
BLO	Quatro blocos, representados pelos números: 1, 2, 3, 4.
CER	(Notas de 0 a 5). Nota 0: plantas que morreram no início do estudo, Nota 1: sem cercóspora; 2 a 5: folhas com cercóspora (2=pouca e 5=muita).
FER	(Notas de 0 a 5). Nota 0: plantas que morreram no início do estudo, 1 = ausência de pústulas e reações de hipersensibilidade; 2 = poucas folhas com pústulas sem esporos (“flecks”) e com reações de hipersensibilidade; 3 = poucas pústulas por folha com alta produção de esporos e pouco distribuídas; 4 = média quantidade de pústulas por folha, distribuídas na planta com alta produção de esporos; 5 = alta quantidade de pústulas com alta produção de esporos e alta desfolha da planta.
BM	(Notas de 0 a 5), Nota 0: plantas que morreram no início do estudo, 1 = Imune, folhas sem nenhuma lesão; 2 = Folhas com poucas lesões de forma afilada; 3 = Folhas com poucas e pequenas lesões; 4 = Folhas com moderada infestação e lesões típicas e com larvas vivas; 5 = Folhas com severa infestação e lesões típicas e com larvas vivas.
TF	(Notas de 0 a 5), Nota 0: plantas que morreram no início do estudo, 1 = miúdo (Ref. ‘Ibairi’), 2 = pequeno, 3 = médio, 4 = grande e 5 = graúdo (Ref. ‘Maragogipe’).
CMT	(Notas de 0 a 5), Nota 0: plantas que morreram no início do estudo, 1 = precoce; 2 = média para precoce; 3 = média; 4 = média para tardia; 5 = tardia.
UMT	(Notas de 0 a 4), Nota 0: plantas que morreram no início do estudo, 1 = uniforme; 2 = medianamente uniforme; 3 = medianamente desuniforme; 4 = desuniforme.

FIL: Fileiras; BLO: Blocos; CER: Incidência de cercosporiose; FER: Incidência de ferrugem; BM: Infestação de Bicho mineiro; TF: Tamanho do fruto; CMT: Ciclo de maturação; UMT: Uniformidade de maturação dos frutos.

No que se refere às covariáveis contínuas, as unidades de medida utilizadas na coleta dos dados estão descritas na Tabela 4. Os dados de temperatura (TEMP) foram obtidos junto ao Instituto Nacional de Meteorologia (INMET, 2025), enquanto os dados de precipitação (PREC) foram fornecidos pelo Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN, 2025).

As demais covariáveis foram registradas durante o estágio de maturação dos frutos, nos períodos de colheita dos anos de 2018, 2020, 2021 e 2022.

Tabela 4 – Covariáveis contínuas.

Abreviação	Níveis
VIG	Notas variando de 1 a 10. Nota 1: Planta totalmente depauperada. Nota 10: planta considerada com vigor máximo.
APL	Em centímetros: medido no ramo ortotrópico principal, da superfície do solo ao ponto final de crescimento do ramo.
DCO	Em centímetros: medido no sentido transversal à linha de plantio, medindo-se a maior projeção da “saia” do cafeeiro.
DCA	Em milímetros: medido com auxílio de um paquímetro digital, na região do coleto da planta (+ ou – 5 cm da superfície do solo).
NPR	Quantidade de pares de ramos plagiotrópicos na haste principal.
CRP	Em centímetros: medição no terço médio.
NNR	Quantidade de nós no ramo plagiotrópico.
PROD	Em litros/planta, estimada subjetivamente (visualmente).
TEMP	Temperatura média anual medida em graus Celsius.
PREC	Precipitação acumulada anual, medida em milímetros cúbicos.

VIG: Vigor vegetativo; APL: Altura das plantas; DCO: Diâmetro da copa; DCA: Diâmetro do caule; NPR: Número de pares de ramos plagiontrópicos na haste principal; CRP: Comprimento de ramo plagiotrópico representativo da planta; NNR: Número de nós no ramo plagiotrópico; PROD: Produção; TEMP: Temperatura; PREC: Precipitação.

## 3.2 Métodos

No presente estudo, a resposta foi medida em anos, e o evento de interesse foi definido por meio da comparação das médias anuais de produção das testemunhas em relação ao tratamento.

No primeiro ano de avaliação (2018), a testemunha Paraíso MG H419-1 apresentou uma produção média de 1 litro por planta. Assim, as plantas da geração F2 que atingiram ou excederam esse valor foram consideradas como tendo alcançado o evento de interesse, enquanto as demais permaneceram em observação para possível ocorrência do evento nos anos seguintes. Em 2020, a mesma testemunha registrou uma produção média de 2,86 litros por planta, sendo que, nesse ano, as plantas que produziram igual ou acima desse valor foram consideradas como falha, enquanto as demais continuaram sob risco. Nos anos subsequentes, as médias de produção foram de 1,17 litros por planta em 2021 e 0,725 litros por planta em 2022.

Para a testemunha Catuaí Vermelho, as médias de produtividade observadas foram: 0,475 litros por planta em 2018, 3,2625 litros por planta em 2020, 0,47 litros por planta em 2021 e 0,725 litros por planta em 2022. Esses valores serviram como parâmetro para a definição do evento de interesse em cada ano, seguindo o mesmo critério de comparação aplicado à testemunha Paraíso MG H419-1.

Deste modo, o “tempo de falha” é o período que se estende desde o início do estudo até o momento em que a planta apresenta uma produção média igual ou superior as testemunhas em estudo. A coleta de dados foi realizada anualmente, no mês de março dos anos de 2018,

2020, 2021 e 2022, no estágio fenológico de maturação fisiológica dos grãos. Abrangendo 12 características fenotípicas das plantas relacionadas à sanidade, produção e qualidade da bebida. É importante destacar que, no ano de 2019, devido a restrições de financiamento, a coleta de dados não pôde ser realizada.

Para avaliar a probabilidade de produção das plantas até o tempo  $t$ , estimaram-se as funções de Kaplan-Meier e a função de distribuição acumulada. As comparações entre as curvas de produção dos diferentes níveis das variáveis categóricas utilizadas foram realizadas mediante os testes de log-rank (ênfase em diferenças tardias) e Wilcoxon (sensível a diferenças iniciais). Esta abordagem dupla permite detectar tanto variações precoces quanto tardias no desempenho produtivo entre as plantas.

Com base na equação proposta por Chalita, Colosimo e Demétrio (2002), a proporção de empates é definida como:  $pe = \frac{r-k}{n}$ , em que  $r$  representa o número total de falhas,  $k$  o número de falhas distintas e  $n$  o número total de observações. Verificou-se que a proporção de empates ultrapassou 25%, cenário no qual os autores recomendam a adoção de modelos discretos para a modelagem dos dados. No entanto, modelos que empregam aproximações da função de verossimilhança parcial, como os propostos por Breslow, Efron e exata, podem fornecer resultados satisfatórios e, por essa razão, foram adotados no presente estudo. O cumprimento dos pressupostos de riscos proporcionais foi testado utilizando o teste de Schoenfeld como procedimento de validação dos modelos.

Dessa forma, a análise foi realizada em duas etapas: inicialmente, utilizando os modelos que incorporam as aproximações de Breslow, Efron e exata, respectivamente; e, posteriormente, com os modelos discretos de Cox e logístico. Chalita (1997) e Colosimo *et al.* (2000) indicaram que o teste *score* evidencia a equivalência entre os modelos logístico e de Cox.

Neste estudo, ambos os modelos foram empregados, sendo selecionado aquele que apresentou os menores valores dos critérios de informação AIC e BIC. A validação dos modelos selecionados foi realizada por meio da curva ROC.

As análises foram realizadas utilizando o *software* R (R, 2025). Um resumo dos dados brutos encontra-se sumarizado no Apêndice A, e as rotinas computacionais estão disponíveis no Apêndice B. Os *scripts* foram adaptados de Colosimo e Giolo (2006) e Giolo (2023). O processo de seleção de covariáveis, cuja rotina encontra-se descrita no Apêndice C, foi aplicado conforme apresentado na Seção 2.10. As imagens geradas e seus respectivos *scripts* estão disponíveis no Apêndice D.

Adicionalmente, os seguintes pacotes foram utilizados neste estudo:

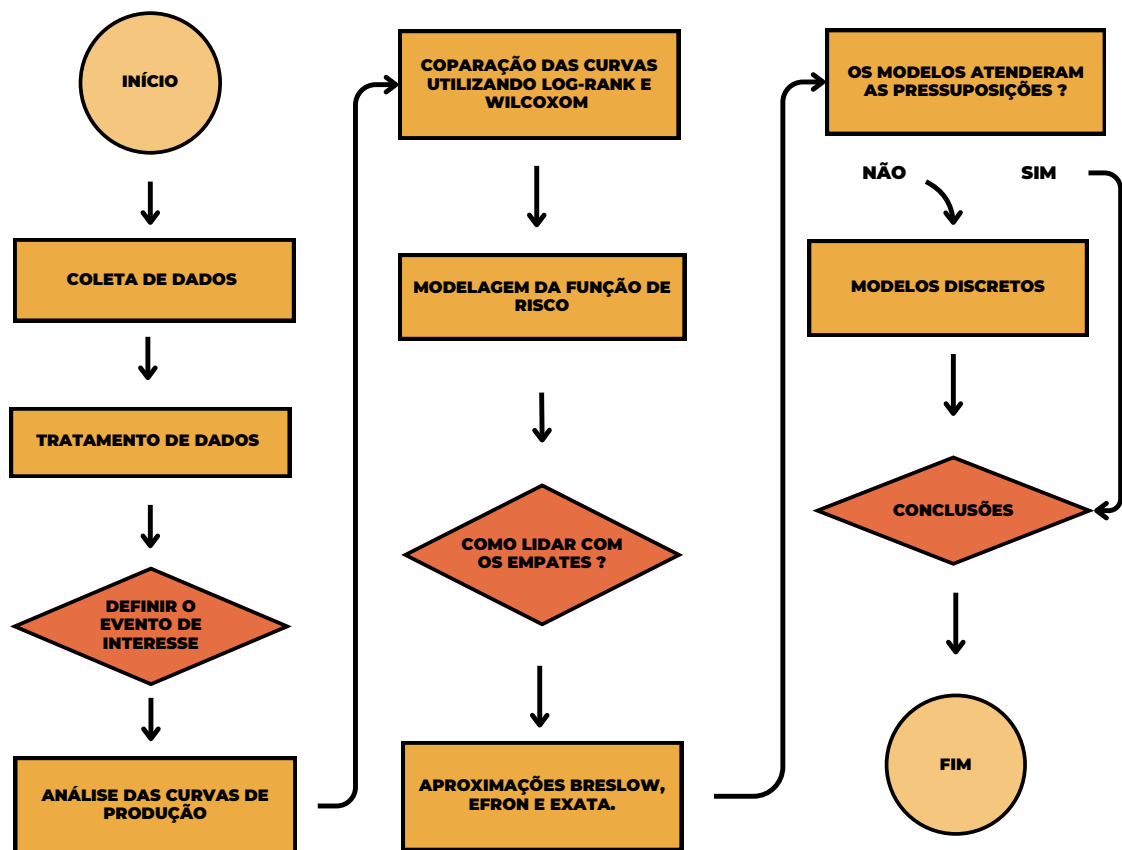
- i) **survival**: Usado com frequência por Colosimo e Giolo (2006), este pacote possibilitou, obter estimativas de Kaplan-Meier (função, *survfit*) e estatísticas do teste de log-rank, pela função *survdiff*.
- ii) **survminer**: Utilizado para criar gráficos de Kaplan-Meier de modo interativo, contendo tabela com número de observações sob risco e também com número de observações censuradas em cada tempo analisado. O pacote permite ainda apresentar no gráfico o resultado do teste log-rank para a covariável analisada. Foi também utilizado para verificar a adequação do modelo de Cox, na análise dos resíduos Schoenfeld.
- iii) **MASS**: Utilizado para ajustar o processo de seleção de covariáveis *stepwise*.

iv) O pacote pROC, foi utilizado para plotar a curva ROC.

Os resultados e discussões serão divididos em dois capítulos: o primeiro abordará a comparação do tratamento com a testemunha Paraíso MG H419-1, e o segundo, a comparação com a testemunha Catuaí Vermelho IAC 144.

Um fluxograma representando os passos da metodologia é representado na Figura 6:

Figura 6 – Passo a passo da metodologia utilizada.



Fonte: Autor.

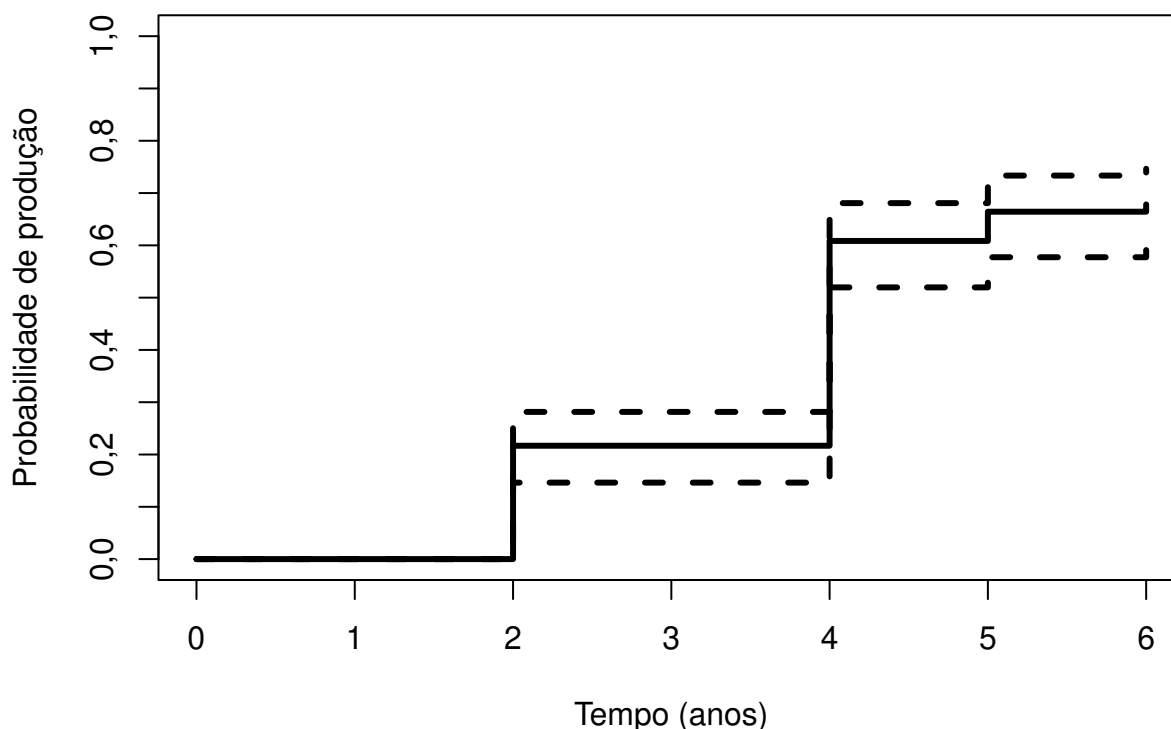
## 4 RESULTADOS E DISCUSSÃO

### 4.1 Paraíso

Em comparação a testemunha Paraíso, observou-se que na população de 200 plantas analisadas, 92 (46%) apresentaram produtividade média igual ou superior à testemunha. A análise de sobrevivência revelou um tempo mediano de ocorrência dos eventos igual a 4, com maior frequência de eventos registrada no ano de 2020.

A Figura 7 representa a curva de produtividade obtida pelo estimador de Kaplan-Meier juntamente com os respectivos intervalos de confiança.

Figura 7 – Curva de produção na comparação com a testemunha Paraíso.



Fonte: Autor.

A Tabela 5 apresenta um resumo detalhado dos tempos de produtividade das plantas obtidas pelo estimador de Kaplan-Meier juntamente com os intervalos de confiança (*IC*) de 95%.

Tabela 5 – Produtividade estimada via Kaplan-Meier em comparação com a testemunha Paraíso.

Tempo	n° risco	n° eventos	n° censura	Produtividade	Erro padrão	<i>IC</i> <sub>95%</sub>
0	200	0	57	0,00	0,00	(1,00; 1,00)
2	143	31	0	0,22	0,04	(0,85; 0,71)
4	112	56	0	0,61	0,10	(0,48; 0,31)
5	56	8	0	0,67	0,11	(0,42; 0,26)
6	48	2	46	0,68	0,12	(0,40; 0,25)

Como complemento à Figura 7, a Tabela 5 apresenta as observações censuradas nos tempos 0 e 6. As plantas censuradas no tempo 0 correspondem àquelas que morreram no início do período de plantio, enquanto as censuradas no tempo 6 referem-se às plantas que não alcançaram produtividade média igual ou superior à da testemunha Paraíso durante os anos de acompanhamento.

No tempo 2, referente ao ano de 2018, registraram-se 31 ocorrências do evento de interesse, correspondendo a uma probabilidade de 22% de ocorrência nesse ano. Nesse período, 143 plantas permaneceram sob risco, ou seja, ainda poderiam ser censuradas ou apresentar produtividade média igual ou superior a testemunha paraíso nos anos subsequentes.

O tempo 4, referente ao ano de 2020, destacou-se como o período com o maior número de eventos de interesse, totalizando 56 ocorrências, o que resultou em uma probabilidade de 61% de ocorrência do evento. Das 56 plantas remanescentes após esse período, observaram-se 8 ocorrências no tempo 5 (ano de 2021) e apenas 2 no tempo 6 (ano de 2022), enquanto as 48 restantes foram censuradas ao final do estudo.

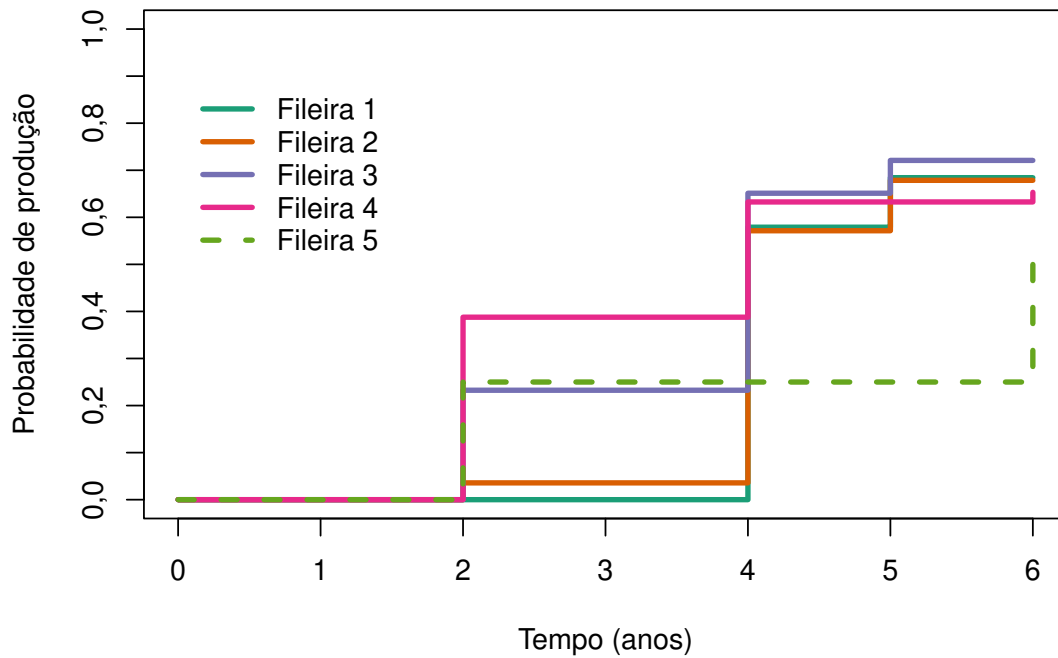
A análise revelou que, do total de plantas monitoradas, 51,5% (103 indivíduos) foram censuradas ao longo do estudo, sendo 57 plantas já no tempo inicial (tempo 0) e outras 46 no momento final do período de acompanhamento (tempo 6). Em contrapartida, 48,5% da população (93 plantas) atingiram produtividade média igual ou superior à testemunha Paraíso, resultando numa probabilidade acumulada de ocorrência deste evento de 68% ao término do período de estudo.

No contexto da experimentação, a quantidade de plantas que atingiram o evento de interesse (92 plantas) indica uma boa capacidade produtiva do cultivar avaliado. Esse desempenho deve ser considerado à luz de diversos fatores: a comparação com um cultivar de alto potencial produtivo previamente testado (Paraíso), o baixo investimento em manejo, a presença de uma população perene e o elevado nível de estresse das plantas, causado por um ambiente propício ao desenvolvimento de doenças, como a ferrugem. A combinação desses fatores, aliada ao número de plantas que apresentaram produtividade igual ou superior à testemunha (Paraíso) ao longo do tempo, sugere um bom potencial para o tratamento avaliado.

#### 4.1.1 Comparações das curvas de produção

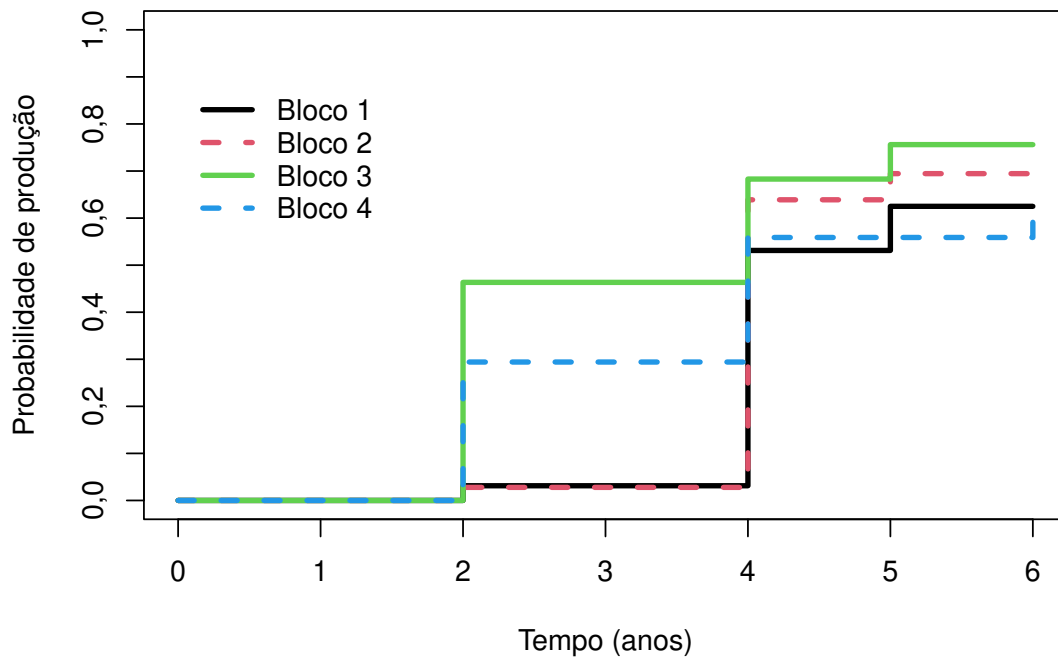
Inicialmente, aplicaram-se os testes globais de log-rank e Wilcoxon com o objetivo de verificar a validade da hipótese nula de igualdade entre as curvas de produção, considerando os diferentes níveis das covariáveis categóricas incluídas no estudo. Apenas as fileiras não apresentaram diferença significativa nas curvas de produtividade, conforme o teste log-rank ( $\chi^2 = 2,3$  e  $\text{valor}_p = 0,7$ ) e o teste de Wilcoxon ( $\chi^2 = 5,5$   $\text{valor}_p = 0,2$ ) indicando que as curvas não diferem em probabilidade de produção entre as cinco fileiras analisadas. A Figura 8 mostra as curvas de produção referentes as fileiras. As demais covariáveis apresentaram diferenças significativas indicadas pelos testes. Um ajuste foi realizado para comparar os diferentes blocos do experimento e a Figura 9 apresenta as diferentes curvas de produtividade.

Figura 8 – Produtividade estimada via Kaplan-Meier para cada nível da covariável fileira: comparação com a testemunha Paraíso.



Fonte: Autor.

Figura 9 – Produtividade estimada via Kaplan-Meier para cada nível da covariável bloco: comparação com a testemunha Paraíso.



Fonte: Autor.

A aplicação do teste de log-rank à variável Bloco não evidenciou diferenças estatisticamente significativas entre as curvas de produtividade (valor- $p = 0,06$ ), considerando um  $\alpha = 5\%$ , indicando ausência de evidências para rejeitar a hipótese nula de igualdade das curvas. Contudo, o teste de Wilcoxon, mais sensível a diferenças iniciais, detectou diferenças significativas (valor- $p = 0,01$ ), sugerindo uma variação no período inicial de produtividade entre os blocos. Diante deste resultado, realizou-se uma análise comparativa dois a dois utilizando o teste de Wilcoxon com o intuito de verificar quais níveis da variável Bloco diferem entre si, os resultados encontram-se detalhados na Tabela 6.

Cada bloco analisado continha 50 plantas. O bloco 1 apresentou 20 eventos de interesse, o bloco 2 apresentou 25, o bloco 3 teve 31 e o bloco 4 registrou 21 eventos. A maior ocorrência de eventos observada no bloco 3 pode estar associada à sua localização em uma região mais adensada do campo experimental, situada na área central da lavoura.

Tabela 6 – Teste de Wilcoxon na comparação das curvas de produção entre os níveis da covariável bloco.

Teste de Wilcoxon		
Comparação	$\chi^2$	Valor- $p$
<b>Blocos</b>		
1 <i>versus</i> 2	0,6	0,500
1 <i>versus</i> 3	8,5	<b>0,004</b>
1 <i>versus</i> 4	1,2	0,300
2 <i>versus</i> 3	7,0	<b>0,008</b>
2 <i>versus</i> 4	0,4	0,500
3 <i>versus</i> 4	2,4	0,100

Valores em negrito indicam comparações com diferença significativa entre os grupos (valor- $p < 0,05$ ).

A diferença acentuada no início das curvas, conforme representado na Figura 8 e confirmada pelo teste de Wilcoxon presente na Tabela 6. O teste indicou diferenças significativas entre os blocos 1 e 3, bem como entre os blocos 2 e 3, pode ser atribuída a um período de replantio. No qual os blocos 1 e 3 apresentaram maior incidência de replantas.

Essa diferença também pode ser explicada pelas características desse período, marcado por uma transição nas plantas, que se encontravam em estágios de desenvolvimento, com características produtivas desuniformes. Com o passar do tempo, no entanto, observou-se uma homogeneização da população, uma vez que não foram identificadas diferenças tardias nas curvas de produção.

Para a covariável ferrugem, foram identificados três níveis: notas 0, 1 e 2. A nota 0 refere-se às plantas que morreram no início do estudo. A ocorrência do evento de interesse foi observada apenas nos níveis 1 e 2; os demais níveis não foram registrados, o que sugere elevada resistência das plantas à doença. No nível 1, foram identificadas 51 plantas, das quais 15 apresentaram produtividade média igual ou superior à da testemunha Paraíso. No nível 2, 93 plantas foram classificadas, sendo que 82 atingiram o evento de interesse.

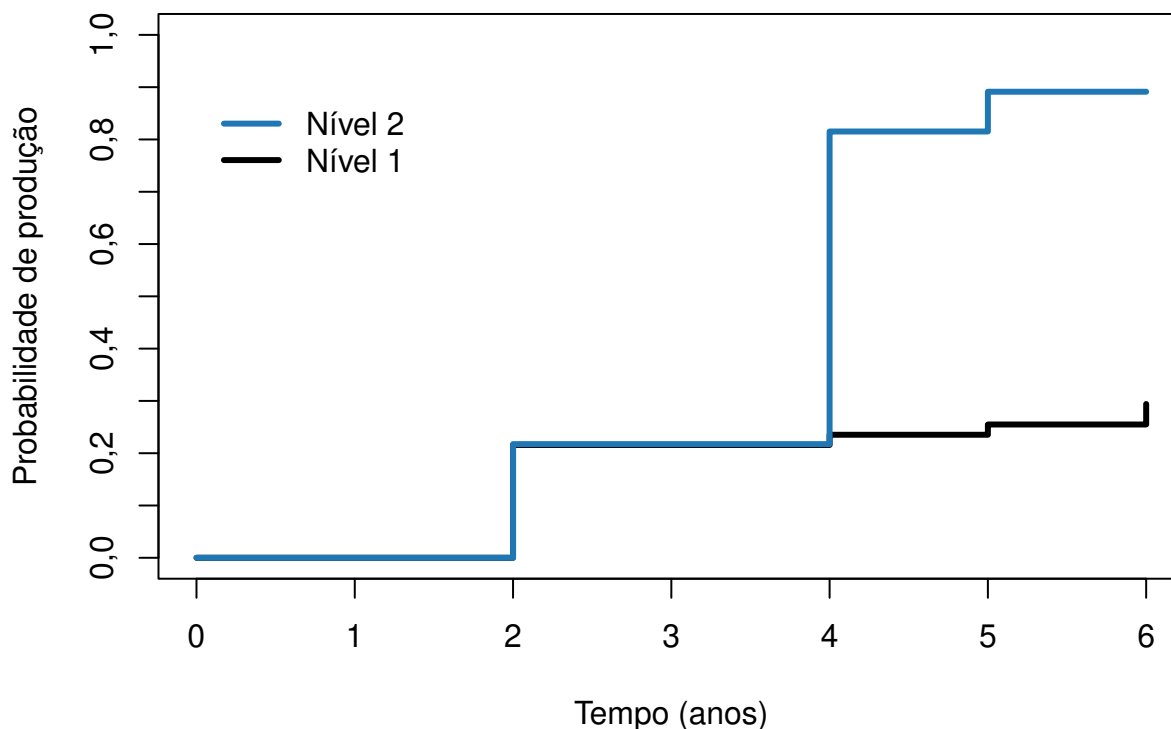
As curvas de produção associadas a cada nível, bem como os resultados dos teste de comparação entre os níveis, são apresentados a seguir na Tabela 7 e na Figura 10.

Tabela 7 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável ferrugem.

Ferrugem	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Nível 1 <i>versus</i> Nível 2	38,8	<b>&lt;0,0001</b>	27,9	<b>&lt;0,0001</b>

Valores em negrito indicam que os grupos diferem entre si (valor – *p* < 0,05).

Figura 10 – Produtividade estimada via Kaplan-Meier para cada nível da covariável ferrugem: comparação com a testemunha Paraíso.



Fonte: Autor.

Os resultados obtidos indicaram um cenário promissor, especialmente diante do elevado número de plantas que demonstraram resistência à doença (Notas 1 e 2). A ausência de identificação de plantas com o evento de interesse em associação com níveis elevados da doença reforça a hipótese de que a produtividade está relacionada à redução dos danos causados por patógenos.

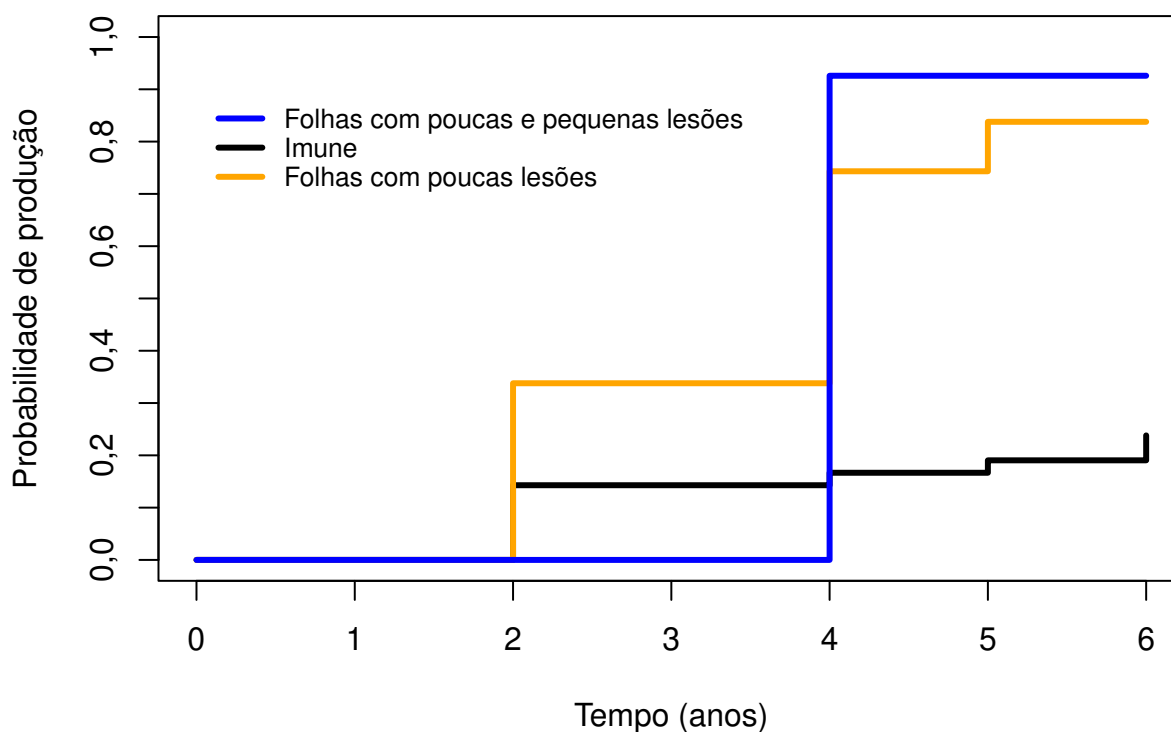
Um fator que corrobora esses achados é o estudo de Mariz *et al.* (2025), que obteve sucesso na piramidação de genes de resistência na população analisada, evidenciando que o objetivo de desenvolver uma população com resistência à ferrugem foi alcançado.

O bicho mineiro é uma doença que provoca senescência foliar prematura, um processo caracterizado pela morte antecipada das folhas, resultando na redução da capacidade fotos-

sintética e, conseqüentemente, na diminuição da produtividade, conforme descrito por Pereira *et al.* 2003.

Para a covariável em questão, 56 plantas apresentaram nota 0, referente as plantas que morreram no início do estudo, 42 apresentaram notas 1, sendo que 10 dessas plantas, alcançaram o evento de interesse. 75 plantas apresentaram nota 2, das quais 62 apresentaram o evento, ainda tiveram 27 plantas que apresentaram nota 3, com 25 apresentando o evento de interesse.

Figura 11 – Produtividade estimada via Kaplan-Meier para cada nível da covariável bicho mineiro.



Fonte: Autor.

Tabela 8 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável bicho-mineiro.

Comparação	Teste de Log-rank		Teste de Wilcoxon	
	$\chi^2$	Valor- <i>p</i>	$\chi^2$	Valor- <i>p</i>
Nível 1 <i>versus</i> Nível 2	36,4	<b>&lt;0,0001</b>	30,3	<b>&lt;0,0001</b>
Nível 1 <i>versus</i> Nível 3	24,4	<b>&lt;0,0001</b>	23,0	<b>&lt;0,0001</b>
Nível 2 <i>versus</i> Nível 3	0,5	0,5000	1,3	0,3000

Valores em negrito indicam que os grupos diferem entre si (valor-*p* < 0,05). Nível 1: plantas imunes à doença; Nível 2: plantas com poucas lesões; Nível 3: plantas com poucas e pequenas lesões.

Ambos os testes, log-rank e Wilcoxon indicaram diferenças nas curvas de produção referentes as notas da covariável, os resultados referentes as comparações 2 a 2 e o valor da estatística dos testes de comparação aplicados na covariável Bicho mineiro (BM) são apresentados na Tabela 8 e na Figura 11.

Os resultados obtidos foram promissores, uma vez que foram identificados exclusivamente plantas classificadas nos níveis 1, 2 e 3, os quais estão associados a indivíduos mais resistentes ou com poucas e pequenas lesões foliares.

A proximidade das curvas referentes aos níveis 2 (plantas com poucas lesões) e 3 (plantas com poucas e pequenas lesões), conforme ilustrado na Figura 11 e corroborado pelos resultados apresentados na Tabela 8, foi confirmada pelos testes de log-rank e Wilcoxon. Esse achado reforça o potencial do tratamento, que não apresentou prejuízos quanto à produtividade nesses níveis.

O nível 1 (plantas resistentes) apresentou baixa produtividade ao longo dos anos. No entanto, destaca-se que as plantas deste nível estiveram associadas a um maior número de recorrências do evento ao longo do período de observação: todas as 10 plantas classificadas no nível 1 que apresentaram o evento de interesse tiveram recorrência duas ou mais vezes ao longo do tempo.

Diante desse comportamento, estudos que considerem eventos recorrentes podem constituir uma alternativa metodológica interessante para aprofundar a análise deste fenômeno e avaliar suas consequências sobre a produtividade.

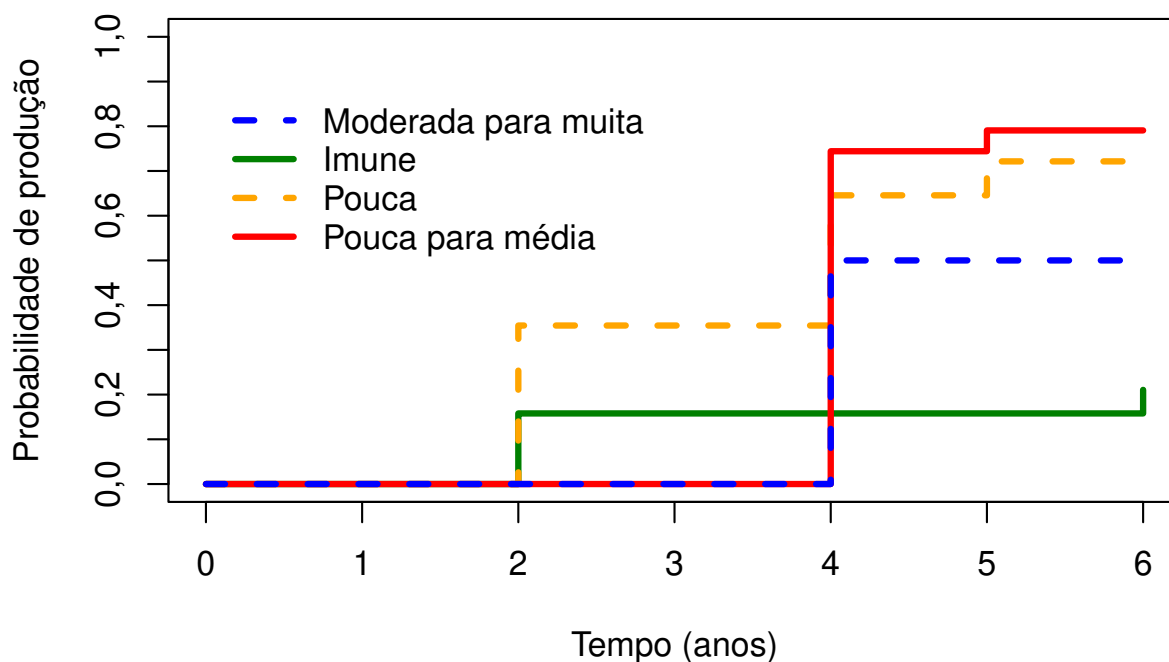
Para covariável cercosporiose, 56 plantas apresentaram nota 0, referente as plantas que morreram no início do estudo, 19 plantas apresentaram notas 1, sendo que 4 dessas plantas, alcançaram o evento de interesse. 79 plantas apresentaram nota 2, das quais 58 apresentaram o evento, ainda tiveram 44 plantas que apresentaram nota 3, das quais 34 apresentaram o evento. Duas plantas apresentaram o nível 4, e 1 delas apresentou o evento de interesse.

Tabela 9 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável cercosporiose.

Comparação	Teste de Log-rank		Teste de Wilcoxon	
	$\chi^2$	Valor- <i>p</i>	$\chi^2$	Valor- <i>p</i>
Nível 1 <i>versus</i> Nível 2	14,1	<b>0,0002</b>	11,9	<b>0,0006</b>
Nível 1 <i>versus</i> Nível 3	12,9	<b>0,0003</b>	12,6	<b>0,0004</b>
Nível 1 <i>versus</i> Nível 4	0,8	0,5000	0,5	0,5000
Nível 2 <i>versus</i> Nível 3	1,0	0,3000	2,7	0,1000
Nível 2 <i>versus</i> Nível 4	0,6	0,4000	0,8	0,4000
Nível 3 <i>versus</i> Nível 4	0,8	0,4000	0,7	0,4000

Valores em negrito indicam que os grupos diferem entre si (valor-*p* < 0,05). Nível 1: plantas imunes à doença; Nível 2: plantas com poucas lesões; Nível 3: plantas com poucas a médias lesões; Nível 4: plantas com médias a moderadas lesões.

Figura 12 – Produtividade estimada via Kaplan-Meier para cada nível da covariável cercosporiose: comparação com a testemunha Paraíso.



Fonte: Autor.

Ambos os testes, log-rank e Wilcoxon indicaram diferenças nas curvas de produção referentes as notas da covariável, os resultados referentes as comparações 2 a 2 e o valor da estatística dos testes de comparação aplicados na covariável cercosporiose (CER) são apresentados na Tabela 9, bem como a Figura 12 que apresenta as curvas de produtividade para cada nível. A curva de produção da nota 1 diferiu das curvas das notas 2 e 3, conforme indicado na Tabela 9. As demais curvas não apresentaram diferenças significativas de produtividade em nenhum dos testes realizados.

De acordo com Lima *et al.* 2018, práticas adequadas de manejo podem reduzir o impacto da cercosporiose. Assim, a elevada incidência da doença observada neste estudo pode estar associada ao baixo investimento no programa de melhoramento, resultando em um manejo menos eficiente e em menor vigor das plantas, o que, conseqüentemente, favorece uma maior incidência da enfermidade.

Esse cenário corrobora a elevada quantidade de plantas com notas associadas à incidência da doença. Destaca-se que, mesmo sob essas condições, as plantas que apresentaram pouca ou pouca a média incidência de cercosporiose alcançaram uma produtividade próxima de 80%, evidenciando o potencial produtivo do cultivar.

Ao se avaliar a resistência do cultivar às doenças de forma geral, verificou-se que a maioria das plantas que apresentaram o evento de interesse possuía índices classificados com as notas 2 e 3, correspondentes a folhas com poucas e pequenas lesões. Os resultados indicam que plantas com baixa ou moderada incidência da doença apresentaram produtividade compatível com a testemunha Paraíso, reforçando a relação entre menor gravidade da doença e desempenho produtivo satisfatório. Esses achados estão em conformidade com os relatados por Poma *et al.* (2024) e Avelino *et al.* (2015).

Para avaliação da qualidade da bebida, consideraram-se três características principais:

uniformidade dos grãos, ciclo de maturação e tamanho do fruto.

Tabela 10 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável uniformidade de maturação dos frutos.

Comparação	Teste de Log-rank		Teste de Wilcoxon	
	$\chi^2$	Valor- $p$	$\chi^2$	Valor- $p$
Nota 1 <i>versus</i> Nota 2	5,3	<b>0,0200</b>	5,0	<b>0,0200</b>
Nota 1 <i>versus</i> Nota 3	13,1	<b>0,0003</b>	11,6	<b>0,0007</b>
Nota 1 <i>versus</i> Nota 4	2,3	0,1000	2,3	0,1000
Nota 2 <i>versus</i> Nota 3	5,9	<b>0,0200</b>	2,6	0,1000
Nota 2 <i>versus</i> Nota 4	3,5	0,0600	3,9	0,0500
Nota 3 <i>versus</i> Nota 4	15,0	<b>0,0001</b>	12,3	<b>0,0005</b>

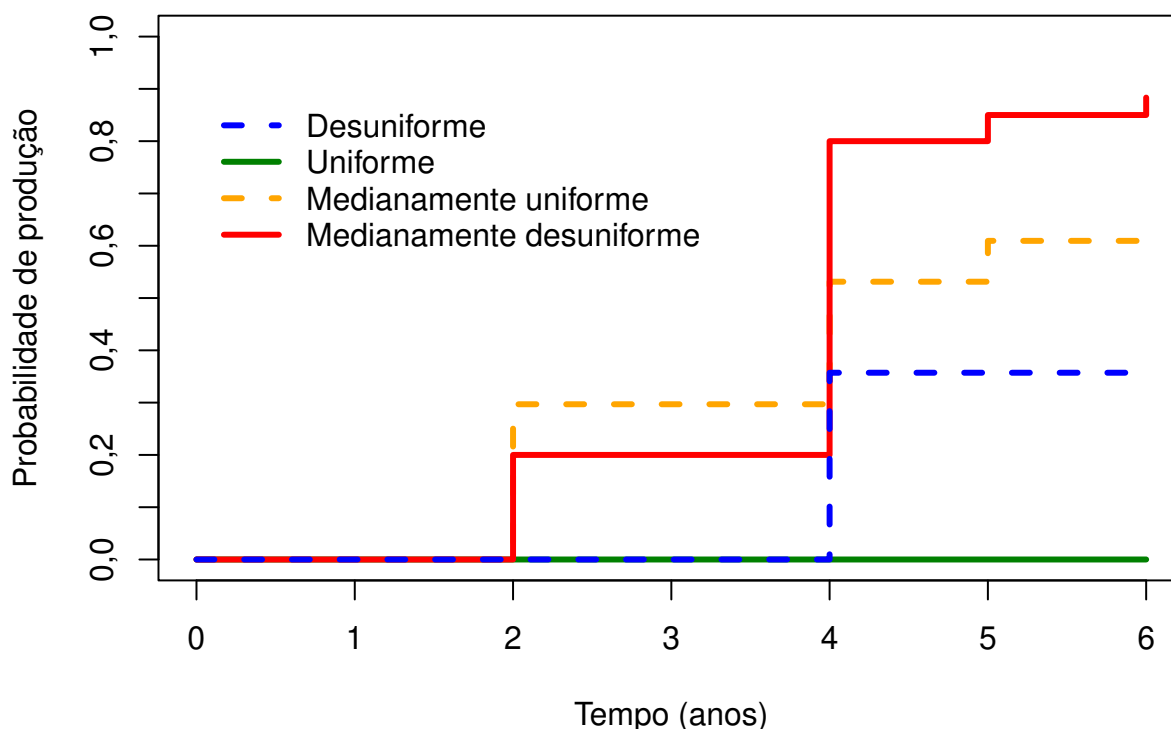
Valores em negrito indicam que os grupos diferem entre si (valor- $p < 0,05$ ). Nota 1: uniforme; Nota 2: medianamente uniforme; Nota 3: medianamente desuniforme; Nota 4: desuniforme.

Os testes de log-rank e Wilcoxon indicaram diferenças significativas entre as curvas de produtividade nos diferentes níveis de uniformidade de maturação. O teste log-rank apresentou estatística  $\chi^2 = 21,8$  (Valor- $p < 0,0001$ ), enquanto o teste Wilcoxon mostrou  $\chi^2 = 17,4$  (Valor- $p = 0,0006$ ).

Na análise, identificaram-se cinco níveis da covariável. O nível 0 correspondeu a 56 plantas que não sobreviveram ao início do plantio. No nível 1, registraram-se 5 plantas, nenhuma das quais atingiu o evento de interesse. Para o nível 2, observaram-se 65 plantas, sendo que 39 apresentaram o evento de interesse. O nível 3 compreendeu 60 plantas, das quais 53 demonstraram produtividade média igual ou superior à testemunha 'Paraíso'. Por fim, no nível 4, contabilizaram-se 14 plantas, com 5 apresentando o evento de interesse. Procedeu-se com a comparação 2 a 2 para verificar quais níveis diferem entre si, os valores da estatística do teste  $\chi^2$  e Valor- $p$  associado as comparações, se encontram na Tabela 10. Um gráfico com as curvas de produtividade para cada nível se encontram na Figura 13.

Os resultados revelaram diferenças significativas nos padrões de maturação entre os grupos avaliados. As plantas classificadas com nota 1, correspondentes à maturação uniforme, demonstraram diferenças estatisticamente significativas em relação às notas 2 (maturação medianamente uniforme) e 3 (maturação medianamente desuniforme).

Figura 13 – Produtividade estimada via Kaplan-Meier para cada nível da covariável uniformidade de maturação dos frutos.



Fonte: Autor.

Contudo, não foram observadas diferenças significativas entre as curvas das notas 1 (Uniformes) e 4 (plantas desuniforme) em nenhum dos dois testes aplicados, conforme indicado na Tabela 10.

Os grupos classificados como notas 2 e 3, que representaram as categorias com maior número de plantas, apresentaram resultados divergentes. O teste log-Rank detectou diferenças significativas entre esses grupos ( $\chi^2 = 5,9$  valor  $-p = 0,02$ ), enquanto o teste Wilcoxon não rejeitou a hipótese de igualdade entre as curvas ( $\chi^2 = 2,6$  valor  $-p = 0,1$ ). Esta discrepância é consistente com as características metodológicas de cada teste, particularmente pelo fato do teste Wilcoxon atribuir maior peso aos eventos iniciais, conforme evidenciado na Figura 13, onde se observa que as curvas de produtividade começam a divergir apenas na fase final do período de acompanhamento, após manterem-se próximas durante a fase inicial.

Para a covariável tamanho do fruto ambos os testes também indicaram diferenças entre as curvas de produção. O teste de teste log-Rank apresentou estatística  $\chi^2 = 14,9$  (Valor  $-p = 0,0006$ ), enquanto o teste Wilcoxon mostrou  $\chi^2 = 17,6$  (Valor  $-p = 0,0004$ ).

Atribuiu-se a nota 0 às 56 plantas que não sobreviveram à fase inicial do experimento. Não foram observadas atribuições das notas 1 e 5 em nenhuma das plantas avaliadas. A ausência da nota 5 deve-se à inexistência de parentais com essa característica entre os materiais utilizados no tratamento. No grupo classificado com nota 2 (pequeno), registraram-se 23 plantas, das quais 8 (34,7%) atingiram o padrão de produtividade igual ou superior a testemunha paraíso.

Tabela 11 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável tamanho dos frutos.

<b>Tamanho do fruto</b>	<b>Teste de log-rank</b>		<b>Teste de Wilcoxon</b>	
	$\chi^2$	<i>p</i> -valor	$\chi^2$	<i>p</i> -valor
Nota 2 <i>versus</i> Nota 3	12,3	< <b>0,0001</b>	15,2	< <b>0,0001</b>
Nota 2 <i>versus</i> Nota 4	17,6	< <b>0,0001</b>	18,9	< <b>0,0001</b>
Nota 3 <i>versus</i> Nota 4	0,5	0,5000	0,1	0,8000

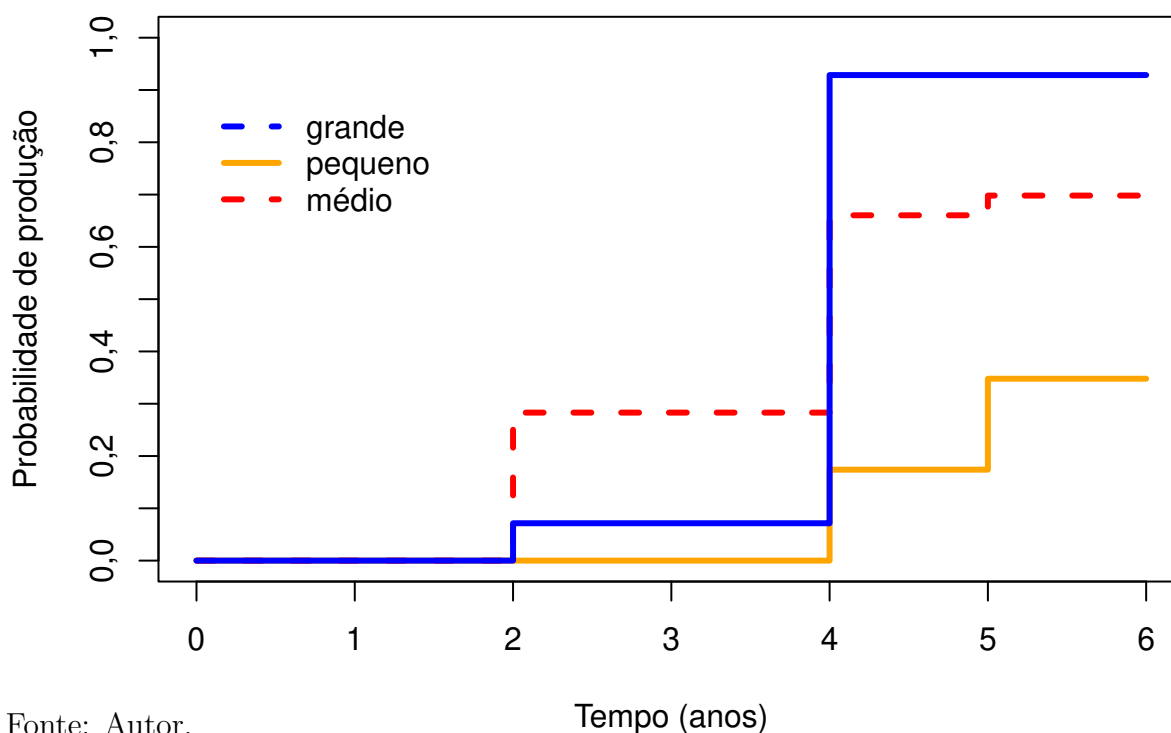
Valores em negrito indicam diferença estatisticamente significativa entre os grupos ( $p < 0,05$ ). Classificação do tamanho dos frutos: Nota 2 (pequeno), Nota 3 (médio), Nota 4 (grande).

A maior concentração de indivíduos foi observada na classe correspondente à nota 3 (porte médio), totalizando 107 plantas, sendo que 76 (71,0%) alcançaram o evento de interesse. Essa elevada frequência de indivíduos com nota 3 está, em grande medida, associada à presença de genitores do tipo híbrido de Timor empregados nos cruzamentos. Por fim, no grupo de plantas designados com nota 4 (grande), identificaram-se 14 plantas, das quais 13 (92,9%) apresentaram produtividade igual ou superior à da testemunha Paraíso.

Foram realizadas comparação 2 a 2 entre os níveis de 2 a 4 com o intuito de verificar a existência de diferenças entre os respectivos níveis. Os resultados dos testes  $\chi^2$  e seus respectivos valores – *p* se encontram na Tabela 11.

As curvas referentes às notas 3 (médio) e 4 (grande) não apresentaram diferenças em nenhum dos testes realizados, enquanto a curva associada ao nível 2 (pequeno) diferiu das curvas 3 e 4, respectivamente como observado na Tabela 11.

Figura 14 – Produtividade estimada via Kaplan-Meier para cada nível da covariável tamanho do fruto: comparação com a testemunha Paraíso.



Fonte: Autor.

Graficamente, podemos observar pela Figura 14 que as curvas referentes aos níveis 3 (médio) e 4 (grande) possuem comportamentos próximos ao longo do tempo, enquanto a curva 2 apresenta diferenças em relação às demais nos últimos dois anos de observação, comportamento que corrobora com os resultados dos testes estatísticos.

Para covariável ciclo de maturação (CMT) os testes de log-rank e Wilcoxon rejeitaram a hipótese de igualdade entre as curvas dos diferentes níveis da covariável. O teste log-rank apresentou estatística  $\chi^2 = 29,2$  (Valor  $-p < 0,0001$ ), enquanto o teste Wilcoxon mostrou  $\chi^2 = 25,4$  (Valor  $-p < 0,0001$ ). No total, foram atribuídas seis notas à covariável em análise.

Verificou-se que 56 plantas receberam a nota 0, correspondente às plantas que morreram no início do estudo, não sendo registrado nenhum caso do evento de interesse. A nota 1 (precoce) foi atribuída a 4 plantas, das quais 25% (uma planta) apresentaram produção média igual ou superior à testemunha Paraíso. A nota 2 (média para precoce) foi observada em 37 plantas, sendo que 54,05% delas (20 plantas) atingiram o evento de interesse. A nota 3 (maturação média) foi a mais frequente, atribuída a 84 plantas, com ocorrência do evento de interesse em 84,52% dos casos (71 plantas). Em relação à nota 4 (média para tardia), 8 plantas foram classificadas nesse nível, com ocorrência do evento em 62,5% delas (5 plantas). Por fim, a nota 5 (tardia) foi atribuída a 11 plantas, não sendo registrada a ocorrência do evento de interesse em nenhuma delas.

Foram realizados testes 2 a 2 para os níveis de 1 a 5 da covariável ciclo de maturação, na Tabela 12 podemos encontrar os testes  $\chi^2$  e seus respectivos valores  $-p$ , na Figura 15 temos as respectivas curvas de produção, associadas a cada nível da covariável.

Tabela 12 – Testes de log-rank e Wilcoxon na comparação das curvas de produção entre diferentes níveis da covariável ciclo de maturação.

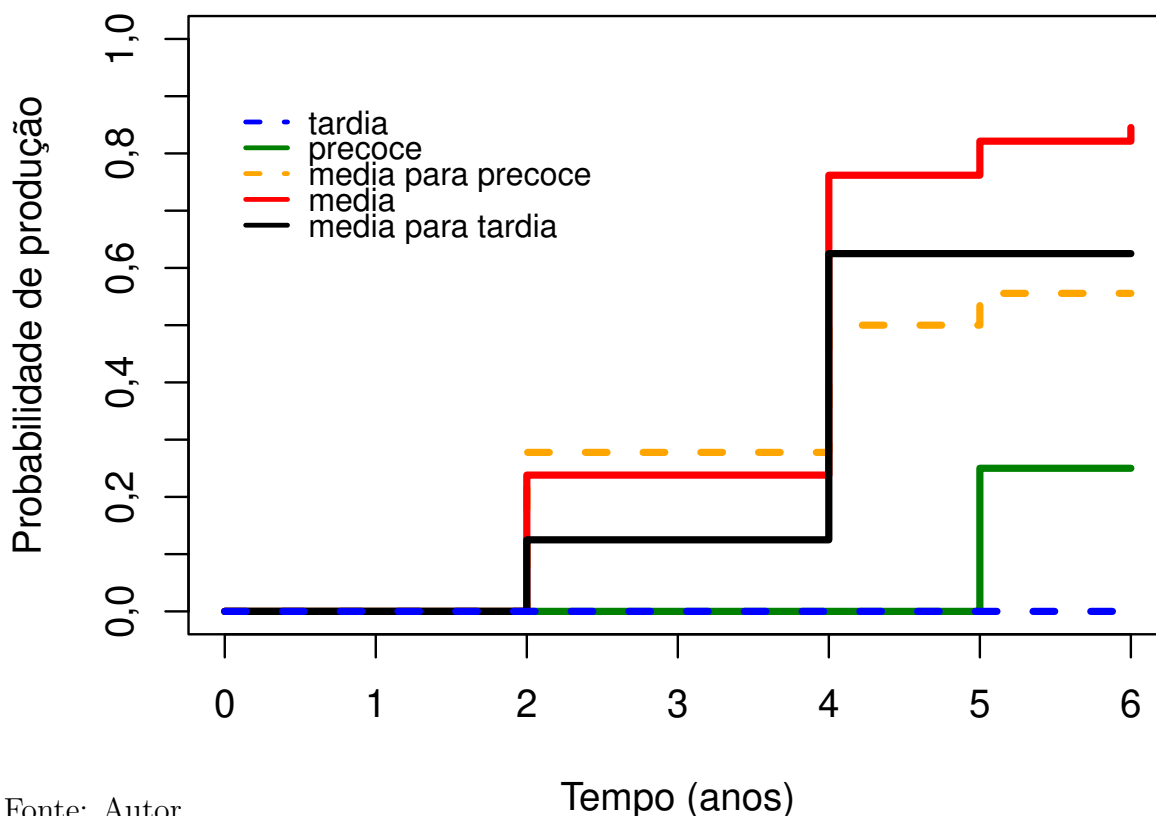
Ciclo de maturação	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	$p$ -value	$\chi^2$	$p$ -value
Nota 1 <i>versus</i> Nota 2	1,4	0,2000	2,0	0,2000
Nota 1 <i>versus</i> Nota 3	6,0	<b>0,0100</b>	7,0	<b>0,0080</b>
Nota 1 <i>versus</i> Nota 4	1,7	0,2000	2,4	0,1000
Nota 1 <i>versus</i> Nota 5	2,8	0,1000	2,8	0,1000
Nota 2 <i>versus</i> Nota 3	6,2	<b>0,0100</b>	3,3	0,0700
Nota 2 <i>versus</i> Nota 4	0,0	0,9000	0,0	1,0000
Nota 2 <i>versus</i> Nota 5	9,3	<b>0,0020</b>	8,9	<b>0,0030</b>
Nota 3 <i>versus</i> Nota 4	1,9	0,2000	1,2	0,3000
Nota 3 <i>versus</i> Nota 5	24,0	<b>&lt; 0,0001</b>	21,0	<b>&lt; 0,0001</b>
Nota 4 <i>versus</i> Nota 5	8,9	<b>0,0030</b>	8,9	<b>0,0030</b>

Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ). Nota 1: precoce; Nota 2: média para precoce; Nota 3: média; Nota 4: média para tardia; Nota 5: tardia.

As curvas referentes as notas 2 (média para precoce), 3 (média) e 4 (média para tardia) não apresentaram diferenças em nenhum dos testes aplicados, assim como ocorreu com as curvas referentes às notas 1 (precoce) e 5 (tardia). Na Tabela 12, observa-se ainda que as plantas de nota 1 não diferiram daquelas de notas 3 e 4, enquanto as plantas de nota 5 diferiram de todos os demais níveis, exceto do nível 1.

A Figura 15 apresenta as curvas de produtividade para os níveis da covariável ciclo de maturação (CMT). Observa-se uma semelhança entre as curvas correspondentes às notas 2, 3 e 4, conforme indicado pelos testes apresentados na Tabela 12, assim como entre as curvas das notas 1 e 5.

Figura 15 – Produtividade estimada via Kaplan-Meier para cada nível da covariável ciclo de maturação: comparação com a testemunha Paraíso.



Fonte: Autor.

A presença de distintos ciclos de maturação configura-se como um achado promissor no que se refere à qualidade da bebida. Essa diversidade fenológica permite o planejamento das atividades de colheita de forma mais estratégica, favorecendo a obtenção de frutos com maturação homogênea em cada ciclo.

Tal variação possibilita a realização de colheitas seletivas, minimizando a presença de grãos imaturos ou subdesenvolvidos no produto final, esses fatores contribuem não apenas para o aprimoramento da qualidade sensorial da bebida, mas também para a redução da demanda por mão de obra e o aumento da eficiência operacional no processo de colheita.

#### 4.1.2 Modelo final utilizando aproximação de Breslow

Conforme o procedimento de seleção de covariáveis descrito na Seção 2.10, o modelo final incorporou as covariáveis Altura da Planta (APL), Diâmetro do Caule (DCA) e Vigor (VIG).

Todas as etapas do processo de seleção de covariáveis estão apresentadas na Tabela 13. Para a avaliação da contribuição individual de cada covariável no modelo, foi elaborada a Tabela 14, com base no Teste da Razão de Verossimilhanças (TRV) para a comparação entre modelos. Nessa tabela, são exibidos a estatística qui-quadrado calculada ( $\chi^2_{\text{cal}}$ ), os respectivos graus de liberdade (gl) e os valores  $-p$  associados a cada teste.

O teste da razão de verossimilhança indicou que todas as covariáveis selecionadas foram estatisticamente significativas para o modelo valor  $-p < 0,05$ , com melhorias consistentes

Tabela 13 – Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Breslow.

PASSO	MODELOS	Looglike	TRV	Valor – <i>p</i>
1	1	NULO	916,0587	-
2		APL	904,8876	11,1711 <b>0,0008</b>
3		BLO	914,9073	1,1514 0,2833
4		BM	899,2401	16,8186 < <b>0,0001</b>
5		CER	913,5370	2,5217 0,1123
6		CRP	914,7948	1,2638 0,2609
7		DCA	814,6857	101,3729 < <b>0,0001</b>
8		DCO	915,7194	0,3393 0,5602
9		FER	887,9835	28,0752 < <b>0,0001</b>
10		FIL	915,5291	0,5296 0,4668
11		NNR	912,4921	3,5666 <b>0,0590</b>
12		NRP	912,4921	1,8677 0,1717
13		VIG	874,7373	41,3213 < <b>0,0001</b>
14		TEMP	791,2307	124,8279 < <b>0,0001</b>
15		PREC	821,0711	94,9876 < <b>0,0001</b>
16	2	APL+BM+DCA+FER+NNR+VIG+TEMP+PREC	749,5338	-
17		BM+DCA+FER+NNR+VIG+TEMP+PREC	755,0390	5,5052 <b>0,0190</b>
18		APL+DCA+FER+NNR+VIG+TEMP+PREC	750,5974	1,0637 0,3024
19		APL+BM+FER+NNR+VIG+TEMP+PREC	751,2125	1,6787 0,1951
20		APL+BM+DCA+NNR+VIG+TEMP+PREC	750,1122	0,5784 0,4469
21		APL+BM+DCA+FER+VIG+TEMP+PREC	752,8833	3,3495 0,0672
22		APL+BM+DCA+FER+NNR+TEMP+PREC	757,4270	7,8932 <b>0,0050</b>
23		APL+BM+DCA+FER+NNR+VIG+PREC	779,9613	30,4275 < <b>0,0001</b>
24		APL+BM+DCA+FER+NNR+VIG+TEMP	749,5586	0,0248 0,8749
25	3	APL+DCA+VIG	783,3866	-
26		APL+DCA+VIG+BM	783,0962	0,2904 0,5899
27		APL+DCA+VIG+FER	782,7446	0,6420 0,4230
28		APL+DCA+VIG+NNR	783,2416	0,1451 0,7033
29		APL+DCA+VIG+TEMP	754,8268	28,5599 < <b>0,0001</b>
30		APL+DCA+VIG+PREC	781,3011	2,0855 0,1487
31	4	APL+DCA+VIG	783,3866	-
32		APL+DCA+VIG+BLO	783,1692	0,2175 0,6410
33		APL+DCA+VIG+CER	782,7284	0,6582 0,4172
34		APL+DCA+VIG+CRP	783,0773	0,3093 0,5781
35		APL+DCA+VIG+DCO	782,7743	0,6124 0,4339
36		APL+DCA+VIG+FIL	782,8605	0,5261 0,4682
37		APL+DCA+VIG+NRP	779,8864	3,5002 0,0614
38	5	APL+DCA+VIG	783,3866	-
39		DCA+VIG	797,4148	14,0281 <b>0,0002</b>
40		APL+VIG	805,5084	22,1217 < <b>0,0001</b>
41		APL+DCA	832,9609	49,5743 < <b>0,0001</b>
	Modelo Final	APL+DCA+VIG	783,3866	-

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor – *p* < 0,05). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Tabela 14 – Teste da razão de verossimilhança para o modelo de aproximações de Breslow.

	Log-verossimilhança	$\chi^2_{\text{cal}}$	gl	Valor- $p$
NULO	-458,03			
APL	-443,72	28,619	1	< <b>0,0001</b>
DCA	-402,75	81,932	1	< <b>0,0001</b>
VIG	-391,69	22,122	1	< <b>0,0001</b>

Valores em negrito indicam que a contribuição da inclusão das covariáveis no modelo é significativa (valor- $p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; DCA: Diâmetro do caule; VIG: Vigor.

nos valores da log-verossimilhança. Esses resultados evidenciam o poder explicativo das covariáveis no ajuste do modelo. Um resumo dos coeficientes estimados, das razões de risco e dos respectivos intervalos de confiança de 95% são apresentados na Tabela 15.

Tabela 15 – Análise dos coeficientes razão de riscos e intervalos de confiança para o modelo de aproximação Breslow.

Variável	Coefficiente	Razão de riscos	Erro padrão	Valor- $p$	$IC_{95\%}$
APL	-0,0170	0,9831	0,0047	<b>0,0002</b>	(0,9741; 0,9922)
DCA	-0,0591	0,9425	0,0140	< <b>0,0001</b>	(0,9170; 0,9688)
VIG	0,6642	1,9431	0,1479	< <b>0,0001</b>	(1,4542; 2,5965)

Valores em negrito indicam rejeição da hipótese nula de que os coeficientes são iguais a zero, com base no teste de Wald (valor- $p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; DCA: Diâmetro do caule; VIG: Vigor.

Podemos destacar que o modelo de riscos proporcionais de Cox indicou que a variável vigor das plantas (VIG) apresentou efeito positivo e significativo sobre a probabilidade de produtividade ser igual ou superior à testemunha Paraíso. Especificamente, o aumento no vigor das plantas elevou em aproximadamente 94% a chance de alcançar produtividade superior à testemunha. Por outro lado, as variáveis altura da planta (APL) e diâmetro do caule (DCA) apresentaram efeitos negativos significativos. O aumento de APL reduziu em cerca de 1,7% o risco da produtividade ser igual ou superior à testemunha, enquanto o aumento de DCA reduziu essa chance em aproximadamente 5,7%.

#### 4.1.3 Adequação do modelo

A Tabela 16 apresenta os resultados do teste de Schoenfeld aplicado às covariáveis do modelo. Verificou-se que a variável VIG atende às pressuposições de riscos proporcionais. No entanto, as demais covariáveis, APL e DCA, assim como o ajuste global do modelo, apresentaram violações dessa pressuposição, evidenciadas pelos valores- $p < 0,05$ , indicando quebra da suposição de riscos proporcionais.

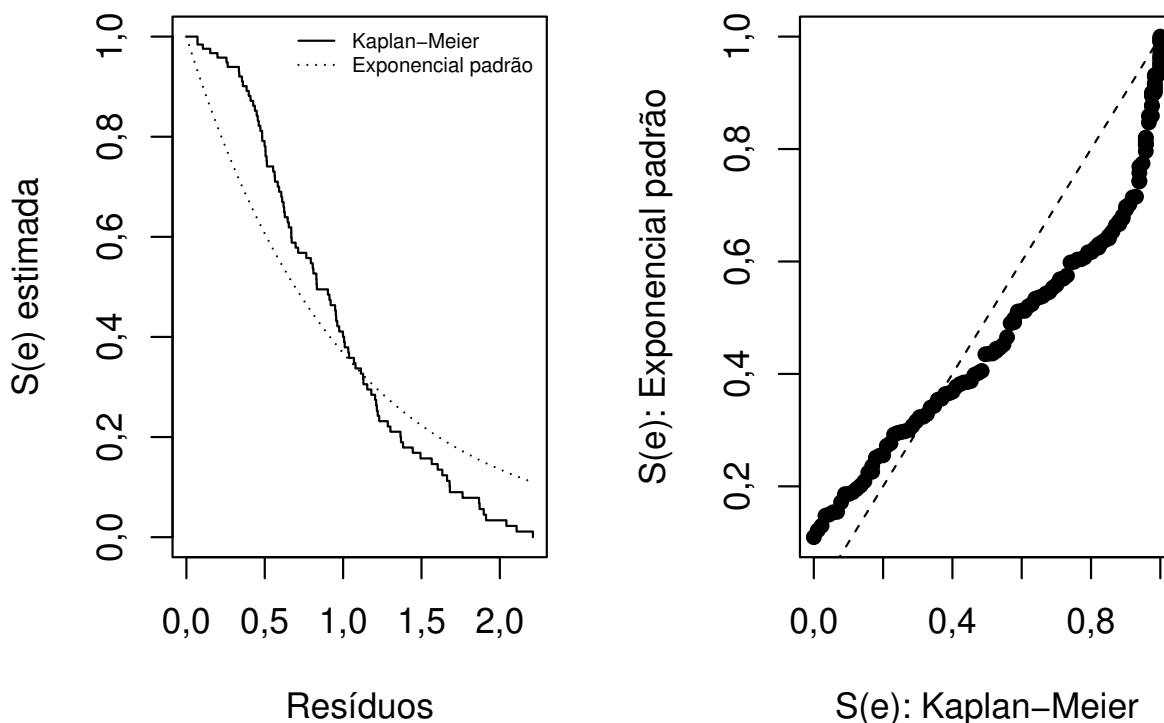
Tabela 16 – Teste de Schoenfeld aplicado as covariáveis do modelo de aproximação de Breslow.

	$\chi^2_{\text{cal}}$	gl	Valor- $p$
APL	18,10	1	< <b>0,0001</b>
DCA	7,65	1	0,0056
VIG	1,57	1	0,2099
GLOBAL	21,26	3	< <b>0,0001</b>

Valores em negrito indicam que as covariáveis não atendem as pressuposições de riscos proporcionais (valor  $-p < 0,05$ ). APL: Altura da planta; DCA: Diâmetro do caule; VIG: Vigor; TEMP: Temperatura; GLOBAL: Ajuste incluindo todas as covariáveis.

A análise dos resíduos com base nos gráficos de Cox-Snell, apresentada na Figura 16, indica que os resíduos não seguem uma distribuição exponencial padrão, o que reforça a inadequação do modelo de Cox aos dados analisados.

Figura 16 – Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação Breslow: Comparação com a testemunha Paraíso.



Fonte: Autor.

A violação das pressuposições de proporcionalidade, aliada ao mau ajuste evidenciado pelos resíduos de Cox-Snell, indica que o modelo ajustado pode não ser adequado. Diante disso, realizou-se um novo ajuste utilizando a aproximação de Efron.

#### 4.1.4 Modelo final utilizando aproximação de Efron

De forma análoga ao procedimento adotado para a aproximação de Breslow, aplicou-se o método de seleção de covariáveis proposto por Collet ao modelo de Cox utilizando a aproximação de Efron. Conforme apresentado na Tabela 17, o processo de seleção resultou em um modelo final contendo as covariáveis altura da planta (APL), diâmetro do caule (DCA), vigor (VIG) e temperatura (TEMP).

Tabela 17 – Seleção de covariáveis usando o modelo de regressão de Cox, com aproximação de Efron.

PASSO	MODELOS	Loglike	TRV	Valor-p	
1	1	NULO	874,270	-	-
2		APL	837,141	37,129	< <b>0,001</b>
3		BLO	873,306	0,964	0,326
4		BM	842,567	31,703	< <b>0,001</b>
5		CER	868,598	5,672	<b>0,017</b>
6		CRP	871,510	2,761	<b>0,097</b>
7		DCA	698,970	175,301	< <b>0,001</b>
8		DCO	873,528	0,742	0,389
9		FER	827,482	46,788	< <b>0,001</b>
10		FIL	873,951	0,319	0,572
11		NNR	866,910	7,360	<b>0,007</b>
12		NRP	866,910	1,299	0,254
13		VIG	816,777	57,493	< <b>0,001</b>
14		TEMP	672,539	201,731	< <b>0,001</b>
15		PREC	707,707	166,563	< <b>0,001</b>
16	2	APL+BM+CER+CRP+DCA+FER+NNR+VIG+TEMP+PREC	596,642	-	-
17		BM+CER+CRP+DCA+FER+NNR+VIG+TEMP+PREC	606,098	9,456	<b>0,002</b>
18		APL+CER+CRP+DCA+FER+NNR+VIG+TEMP+PREC	597,640	0,998	0,318
19		APL+BM+CRP+DCA+FER+NNR+VIG+TEMP+PREC	598,037	1,395	0,238
20		APL+BM+CER+DCA+FER+NNR+VIG+TEMP+PREC	597,948	1,306	0,253
21		APL+BM+CER+CRP+FER+NNR+VIG+TEMP+PREC	599,947	3,305	<b>0,069</b>
22		APL+BM+CER+CRP+DCA+NNR+VIG+TEMP+PREC	597,248	0,606	0,436
23		APL+BM+CER+CRP+DCA+FER+VIG+TEMP+PREC	603,445	6,803	<b>0,009</b>
24		APL+BM+CER+CRP+DCA+FER+NNR+TEMP+PREC	605,089	8,447	<b>0,004</b>
25		APL+BM+CER+CRP+DCA+FER+NNR+VIG+PREC	642,113	45,472	< <b>0,001</b>
26		APL+BM+CER+CRP+DCA+FER+NNR+VIG+TEMP	596,697	0,055	0,814
27	3	APL+DCA+VIG+TEMP	609,486	-	-
28		APL+DCA+VIG+TEMP+BM	607,266	2,220	0,136
29		APL+DCA+VIG+TEMP+CER	606,150	3,336	<b>0,068</b>
30		APL+DCA+VIG+TEMP+CRP	609,480	0,006	0,939
31		APL+DCA+VIG+TEMP+FER	607,840	1,646	0,200
32		APL+DCA+VIG+TEMP+NNR	604,175	5,311	<b>0,021</b>
33	4	APL+DCA+VIG+TEMP	609,486	-	-
34		APL+DCA+VIG+TEMP+BLO	607,015	2,472	0,116
35		APL+DCA+VIG+TEMP+DCO	609,404	0,082	0,775
36		APL+DCA+VIG+TEMP+FIL	604,382	5,104	<b>0,024</b>
37		APL+DCA+VIG+TEMP+NRP	608,586	0,900	0,343
38	5	APL+DCA+VIG+TEMP+NRP	608,586	-	-
39		DCA+VIG+TEMP+NRP	623,928	15,342	< <b>0,001</b>
40		APL+VIG+TEMP+NRP	617,763	9,177	0,002
41		APL+DCA+TEMP+NRP	625,207	16,622	< <b>0,001</b>
42		APL+DCA+VIG+NRP	645,560	36,975	< <b>0,001</b>
43		APL+DCA+VIG+TEMP	609,486	0,900	0,343
44	Modelo Final	APL+DCA+VIG+TEMP	609,486	-	-

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Com o objetivo de avaliar a contribuição individual de cada covariável no modelo, foi elaborada a Tabela 18, com base no Teste da Razão de Verossimilhanças (TRV) para com-

paração entre modelos. A tabela apresenta a estatística qui-quadrado calculada ( $\chi^2_{\text{cal}}$ ), os respectivos graus de liberdade (gl) e os valores- $p$  associados a cada teste.

Tabela 18 – Teste da razão de verossimilhança para o modelo de aproximações de Efron.

	Log-verossimilhança	$\chi^2_{\text{cal}}$	gl	Valor- $p$
NULO	-432,67			
DCA	-349,48	175,301	1	< <b>0,0001</b>
VIG	-337,19	24,596	1	< <b>0,0001</b>
NRP	-324,72	24,936	1	< <b>0,0001</b>
TEMP	-311,96	25,510	1	< <b>0,0001</b>

Valores em negrito indicam que a contribuição da inclusão das covariáveis no modelo é significativa (valor- $p$  < 0,05). NULO: Modelo nulo; DCA: Diâmetro do caule; VIG: Vigor; NRP: Número de ramos plagiontrópicos; TEMP: Temperatura.

Todas as covariáveis analisadas apresentaram valores- $p$  baixos, indicando que são importantes para o modelo. Um resumo dos coeficientes, razão de risco e seus respectivos intervalos de confiança de 95% são apresentados na Tabela 19.

Tabela 19 – Análise dos coeficientes razão de risco e intervalos de confiança, no modelo de Efron.

Covariável	Coefficiente	Razão de riscos	Erro padrão	Valor- $p$	$IC_{95\%}$
DCA	-0,0786	0,9244	0,1220	< <b>0,0001</b>	(0,8909; 0,9591)
VIG	0,6781	1,9703	0,1624	< <b>0,0001</b>	(1,4245; 2,7251)
NRP	-0,0512	0,9501	2,9575	< <b>0,0001</b>	(0,9275; 0,9733)
TEMP	2,4943	12,1136	0,0144	<b>0,0156</b>	(1,6056; 91,3943)

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Wald (valor- $p$  < 0,05). NULO: Modelo nulo; DCA: Diâmetro do caule; VIG: Vigor; NRP: Números de ramos plagiontrópicos; TEMP: Temperatura.

Os preditores DCA, VIG, NRP e TEMP apresentaram valores- $p$  inferiores a 5%, indicando que as covariáveis são estatisticamente diferentes de zero pelo teste Wald e, portanto, importantes para a explicação da produção do cafezal.

A covariável TEMP apresentou um alto valor de risco associado a um aumento de uma unidade na variável. O risco do evento pode variar desde um impacto moderado (razão de risco em torno de 1,60) até um impacto extremo (razão de risco próxima de 91,39). Essa incerteza é, em grande parte, atribuída à baixa variabilidade nos valores da variável TEMP e ao elevado número de empates.

#### 4.1.5 Adequação do modelo

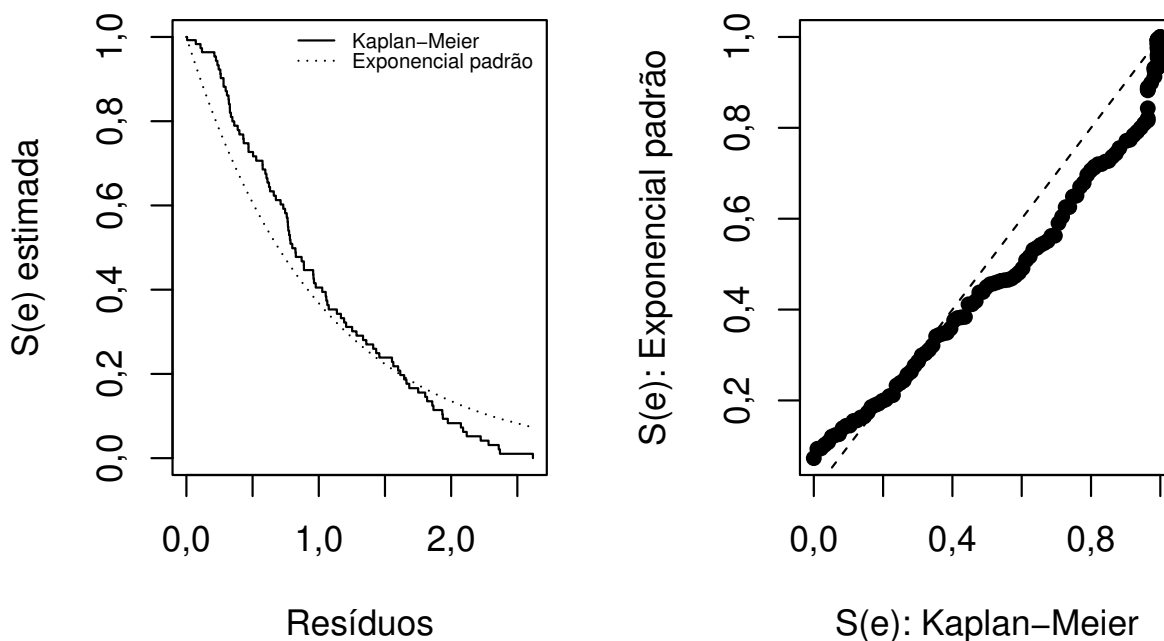
O teste de Schoenfeld aplicado às covariáveis indicou que apenas a covariável VIG não violou as pressuposições do modelo, enquanto as demais apresentaram taxas de falhas não proporcionais. Os resultados do teste para as covariáveis do modelo encontram-se apresentados na Tabela 20.

Tabela 20 – Teste de Schoenfeld aplicado às covariáveis do modelo com aproximação de Efron.

	$\chi^2_{\text{cal}}$	gl	Valor-p
DCA	8,2600	1	<b>0,0041</b>
VIG	0,1540	1	0,6943
NRP	17,0820	1	< <b>0,0001</b>
TEMP	6,5600	1	<b>0,0104</b>
GLOBAL	27,2690	4	< <b>0,0001</b>

Valores em negrito indicam que as covariáveis não atendem às pressuposições de riscos proporcionais (valor  $-p < 0,05$ ). DCA: Diâmetro do caule; VIG: Vigor; NRP: Número de ramos plagiotrópicos; TEMP: Temperatura; GLOBAL: Ajuste incluindo todas as covariáveis.

Figura 17 – Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação Efron: Comparação com a testemunha Paraíso.



Fonte: Autor.

A Figura 17 exibe os resíduos de Cox-Snell obtidos a partir do modelo ajustado com a aproximação de Efron. Os gráficos indicaram um ajuste superior em relação à aproximação de Breslow; contudo, ainda considerado inadequado. Espera-se que os resíduos apresentem um comportamento compatível com uma distribuição exponencial padrão, aproximando-se de uma linha reta com inclinação de  $45^\circ$ . Tal padrão, entretanto, não foi observado nas representações gráficas.

#### 4.1.6 Modelo final utilizando aproximação exact

Após realizado o processo de seleção de covariáveis utilizando a aproximação exata, o modelo final incluiu as covariáveis, NNR (Número de nós nos ramos plagiontrópicos), VIG (Vigor), TEMP (temperatura) e APL (Altura da planta). O passo a passo para o processo

Tabela 21 – Seleção de covariáveis considerando o modelo de regressão de Cox, com aproximação exata.

PASSO	MODELOS	LOGLIKE	TRV	P_VALOR	
1	1	NULO	350,7848	-	-
2		APL	304,6678	46,1170	< <b>0,0001</b>
3		BL	343,3764	7,4083	0,0600
4		BM	304,1740	46,6108	< <b>0,0001</b>
5		CER	331,0542	19,7306	<b>0,0002</b>
6		CRP	348,7977	1,9871	0,1586
7		DCA	183,3966	167,3882	< <b>0,0001</b>
8		DCO	350,2602	0,5245	0,4689
9		FER	309,3473	41,4375	< <b>0,0001</b>
10		FIL	348,4731	2,3117	0,6786
11		NNR	345,1352	5,6496	<b>0,0175</b>
12		NRP	345,1352	2,8987	0,0887
13		VIG	285,8115	64,9733	< <b>0,0001</b>
14		TEMP	156,7782	194,0066	< <b>0,0001</b>
15		PREC	187,2223	163,5624	< <b>0,0001</b>
16	2	APL+BL+BM+CER+DCA+FER+NNR+NRP+VIG+TEMP+PREC	23,7642	-	-
17		BL+BM+CER+DCA+FER+NNR+NRP+VIG+TEMP+PREC	26,5006	2,7363	0,0981
18		APL+BM+CER+DCA+FER+NNR+NRP+VIG+TEMP+PREC	26,2132	2,4490	0,4846
19		APL+BL+CER+DCA+FER+NNR+NRP+VIG+TEMP+PREC	28,4191	4,6549	0,1989
20		APL+BL+BM+DCA+FER+NNR+NRP+VIG+TEMP+PREC	23,9186	0,1544	0,9972
21		APL+BL+BM+CER+FER+NNR+NRP+VIG+TEMP+PREC	23,8512	0,0869	0,7681
22		APL+BL+BM+CER+DCA+NNR+NRP+VIG+TEMP+PREC	23,7996	0,0354	0,9825
23		APL+BL+BM+CER+DCA+FER+NRP+VIG+TEMP+PREC	28,7879	5,0237	<b>0,0250</b>
24		APL+BL+BM+CER+DCA+FER+NNR+VIG+TEMP+PREC	24,1772	0,4130	0,5205
25		APL+BL+BM+CER+DCA+FER+NNR+NRP+TEMP+PREC	31,8629	8,0987	<b>0,0044</b>
26		APL+BL+BM+CER+DCA+FER+NNR+NRP+VIG+PREC	55,0151	31,2509	< <b>0,0001</b>
27		APL+BL+BM+CER+DCA+FER+NNR+NRP+VIG+TEMP	23,6651	0,0992	0,7528
28	3	NNR+VIG+TEMP	85,6249	-	-
29		NNR+VIG+TEMP+APL	54,8196	30,8053	< <b>0,0001</b>
30		NNR+VIG+TEMP+BL	76,9281	8,6968	<b>0,0336</b>
31		NNR+VIG+TEMP+BM	79,5357	6,0892	0,1074
32		NNR+VIG+TEMP+CER	72,3736	13,2513	<b>0,0101</b>
33		NNR+VIG+TEMP+DCA	57,9484	27,6765	< <b>0,0001</b>
34		NNR+VIG+TEMP+FER	79,7647	5,8602	0,0534
35		NNR+VIG+TEMP+NRP	77,5643	8,0606	<b>0,0045</b>
36		NNR+VIG+TEMP+PREC	67,1441	18,4808	< <b>0,0001</b>
37	4	NNR+VIG+TEMP+APL+BL+CER+DCA+NRP+PREC	28,9059	-	-
38		NNR+VIG+TEMP+APL+BL+CER+DCA+NRP+PREC+CRP	28,9004	0,0055	0,9411
39		NNR+VIG+TEMP+APL+BL+CER+DCA+NRP+PREC+DCO	27,6365	1,2694	0,2599
40	5	NNR+VIG+TEMP+APL+BL+CER+DCA+NRP+PREC	28,9059	-	-
41		VIG+TEMP+APL+BL+CER+DCA+NRP+PREC	304,6678	8,0856	<b>0,0045</b>
42		NNR+TEMP+APL+BL+CER+DCA+NRP+PREC	39,6306	10,7248	<b>0,0047</b>
43		NNR+VIG+APL+BL+CER+DCA+NRP+PREC	59,4476	30,5418	< <b>0,0001</b>
44		NNR+VIG+TEMP+BL+CER+DCA+NRP+PREC	30,4118	1,5060	0,2198
45		NNR+VIG+TEMP+APL+CER+DCA+NRP+PREC	29,8904	0,9845	0,8050
46		NNR+VIG+TEMP+APL+BL+DCA+NRP+PREC	32,1849	3,2790	0,5123
47		NNR+VIG+TEMP+APL+BL+CER+NRP+PREC	309,3473	0,3310	0,5651
48		NNR+VIG+TEMP+APL+BL+CER+DCA+PREC	29,4811	0,5752	0,4482
49		NNR+VIG+TEMP+APL+BL+CER+DCA+NRP	30,5888	1,6829	0,1945
50	Modelo Final	NNR + VIG + TEMP + APL	54,8196	-	-

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

de seleção esta disponível na tabela 21

Para a avaliação da contribuição individual de cada covariável no modelo, utilizou-se o teste da razão de verossimilhança, cujos resultados encontra-se na Tabela 22 na qual são exibidas a estatística qui-quadrado calculada ( $\chi_{\text{cal}}^2$ ), os respectivos graus de liberdade (gl) e os valores  $-p$  associados a cada teste.

Tabela 22 – Teste da razão de verossimilhança para o modelo de aproximações exata.

	Log-verossimilhança	$\chi_{\text{cal}}^2$	gl	Valor $-p$
NULO	-175,3920			
NNR	-172,5680	5,6496	1	<b>0,0175</b>
VIG	-142,6150	59,9057	1	<b>&lt; 0,0001</b>
TEMP	-42,8120	199,6046	1	<b>&lt; 0,0001</b>
APL	-27,4100	0,9328	1	<b>&lt; 0,0001</b>

Valores em negrito indicam que a contribuição da inclusão das covariáveis no modelo é significativa (valor  $-p < 0,05$ ). NULO: Modelo nulo; NNR: Número de nós nos ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; APL: Altura da planta.

Todas as covariáveis analisadas apresentaram baixos valores  $-p$ , indicando sua importância para o modelo. Um resumo dos coeficientes, razões de risco e respectivos intervalos de confiança de 95% são apresentados na Tabela 23.

Tabela 23 – Análise dos coeficientes, razão de riscos e intervalos de confiança, no modelo de aproximação exata.

Covariável	Coefficiente	Razão de riscos	Erro padrão	Valor $-p$	$IC_{95\%}$
NNR	-0,1797	0,8355	0,1220	<b>0,0027</b>	(0,7428; 0,9398)
VIG	1,8980	6,6700	0,1624	<b>&lt; 0,0001</b>	(2,6332; 16,8900)
TEMP	16,9000	21.940.000	4,3970	<b>0,0001</b>	(3.969.800; 100.000.000)
APL	-0,0696	12,1136	0,0144	<b>0,0156</b>	(0,9083; 0,9579)

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Wald (valor  $-p < 0,05$ ). NNR: Número de nós nos ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; APL: Altura da planta.

Todas as covariáveis apresentaram valores  $-p$  inferiores a 5%, indicando que diferem significativamente de zero, conforme o Teste de Wald. Observa-se, novamente, que a covariável TEMP apresentou um coeficiente elevado e, conseqüentemente, uma alta razão de riscos. O intervalo de confiança correspondente exibiu um elevado comprimento, o que indica baixa precisão na estimativa do parâmetro.

#### 4.1.7 Adequação do modelo

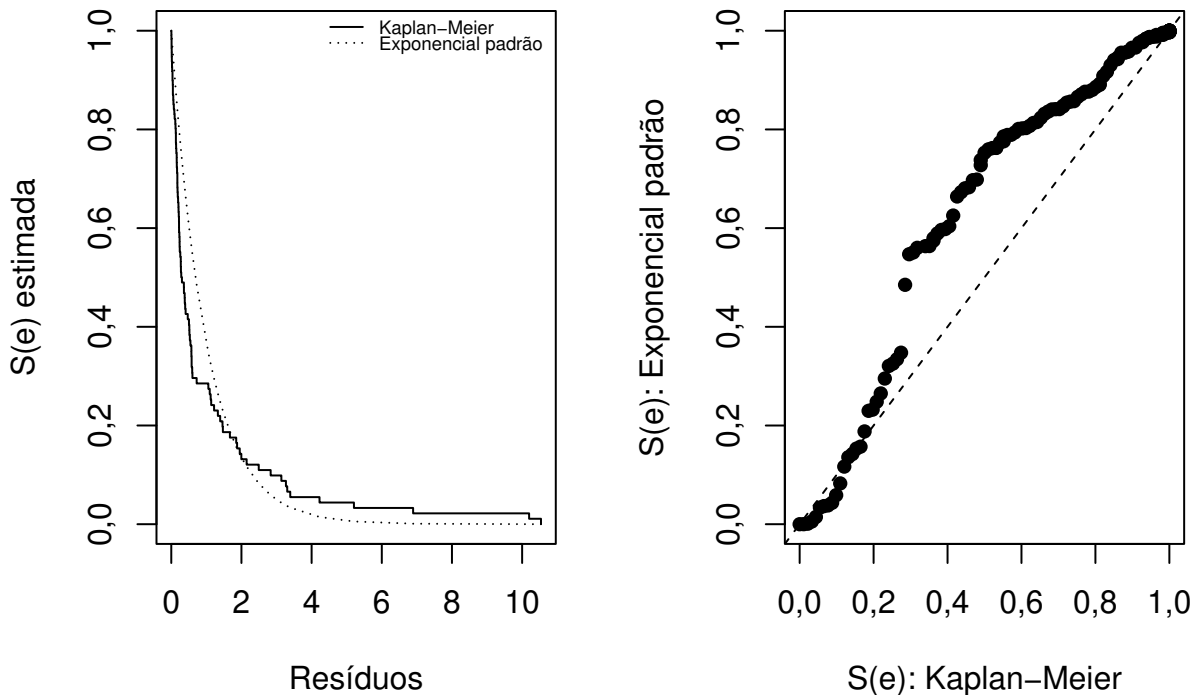
O teste de Schoenfeld aplicado às covariáveis indicou que a covariável NNR e VIG não violaram as pressuposições de riscos proporcionais, enquanto as demais apresentaram taxas de falhas não proporcionais ao longo do tempo. Os resultados do teste para as covariáveis do modelo encontram-se apresentados na Tabela 24.

Tabela 24 – Teste de Schoenfeld aplicado as covariáveis do modelo com aproximação exata.

	$\chi^2_{cal}$	gl	Valor-p
NNR	40,92	1	< <b>0,0001</b>
VIG	1,60	1	0,2100
TEMP	0,71	1	0,4000
APL	30,35	1	< <b>0,0001</b>
GLOBAL	75,24	4	< <b>0,0001</b>

Valores em negrito indicam que as covariáveis não atendem as pressuposições de riscos proporcionais (valor  $p < 0,05$ ). NNR: Número de nós nos ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; APL: Altura da planta.

Figura 18 – Análise dos resíduos de Cox-Snell do modelo ajustado com aproximação exata: Comparação com a testemunha Paraíso.



Fonte: Autor.

A Figura 18 apresenta os resíduos de Cox-Snell obtidos a partir do modelo ajustado com a aproximação 'exact'. Os gráficos indicam um ajuste superior em comparação às demais aproximações avaliadas, embora ainda inadequado. A discrepância entre a curva de Kaplan-Meier e a exponencial padrão, bem como o afastamento dos pontos em relação à reta de 45° no gráfico de probabilidade reforçam a não adequação do modelo ajustado.

#### 4.1.8 Modelos discretos

Devido a baixa qualidade do ajuste para as aproximações para a função de verossimilhança, e os modelos não atenderem as pressuposições de riscos proporcionais de cox a

alternativa proposta na literatura para lidar com dados que possuem muitos empates é a utilização do modelo de regressão Cox para dados agrupados. Nesta etapa atualizou-se o banco de dados e o tempo foi considerado em intervalos, os dados foram organizados como feito em Colosimo e Giolo (2024). O processo de seleção baseou-se na metodologia *stepwise*. Os modelos finais para o modelo discreto de riscos proporcionais e para o modelo logístico são apresentados na Seção 4.1.9.

#### 4.1.9 Modelos finais

Inicialmente, elaborou-se a Tabela 25, que descreve os tempos de produção das plantas analisadas. Percebe-se que todos os intervalos analisados, exceto o primeiro, apresentaram tempos de produção equivalentes, o que corrobora a utilização de modelos discretos.

Tabela 25 – Resumo dos dados apresentados pela tabela de vida.

Intervalo de tempo	Total sob risco	Número de eventos	Número de censuras	$\hat{S}(t)$	Erro padrão de $\hat{S}(t)$
(0, 2]	200	0	57	1,00	–
(2, 4]	143	31	0	0,78	0,0345
(4, 5]	112	56	0	0,39	0,0408
(5, 6]	56	8	0	0,33	0,0395
[6, +∞)	48	2	46	0,32	0,0391

(0, 2]: Intervalo de tempo referente ao período de [2016, 2018]; (2, 4]: Intervalo de tempo referente ao período de [2018, 2020]; (4, 5]: Intervalo de tempo referente ao período de [2020, 2021]; (5, 6]: Intervalo de tempo referente ao período de [2021, 2022]; [6, +∞): Intervalo de tempo referente ao período de [2022, +∞).

O processo de seleção *stepwise* foi realizado com base no critério AIC. O modelo final incluiu os intervalos int1 , int2 int4 juntamente com as covariáveis VIG, TEMP, BM, APL, NNR.

Tabela 26 – Estimativas dos parâmetros dos modelos finais do processo *stepwise* para os modelos de Cox e logístico.

	Modelo de Cox			Modelo logístico		
	Estimativa	Erro padrão	Valor- $p$	Estimativa	Erro padrão	Valor- $p$
$I_1$	-29,9300	18960,0000	0,9987	-29,8848	1507,7408	0,9842
$I_2$	-4,1810	0,6490	< <b>0,0001</b>	-7,7606	1,3181	< <b>0,0001</b>
$I_4$	2,7340	1,6180	0,09107	2,2646	1,5721	0,1497
VIG	0,5573	0,2523	<b>0,0271</b>	1,2217	0,4656	<b>0,0087</b>
APL	-0,0535	0,0104	< <b>0,0001</b>	-0,0800	0,0176	< <b>0,0001</b>
TEMP	5,1550	1,3990	<b>0,0002</b>	5,9149	1,1119	< <b>0,0001</b>
NNR	-0,0927	0,0389	<b>0,0172</b>	-0,2335	0,0725	<b>0,0013</b>

$I_1$ : Intervalo de tempo referente ao período de [2016, 2018];  $I_2$ : Intervalo de tempo referente ao período de [2018, 2020];  $I_4$ : Intervalo de tempo referente ao período de [2021, 2022]; VIG: Vigor vegetativo das plantas; APL: Altura da planta; TEMP: Temperatura; NNR: Número de nós nos ramos plagiontrópicos.

O modelo final selecionado incluiu as covariáveis Vigor vegetativo (VIG), Altura da planta (APL), Temperatura (TEMP), Número de nós nos ramos plagiótropicos (NNR) e os intervalos de tempo  $I_1$  [2016,2018],  $I_2$  [2018,2020] e  $I_4$  [2021,2022], conforme apresentado na Tabela 26.

Entre os modelos analisados, optou-se pelo modelo logístico, em função dos menores valores de AIC (92,4125) e BIC (127,0503), enquanto o modelo discreto de riscos proporcionais de Cox apresentou os valores de AIC (95,1942) e BIC (129,8302).

Observou-se que o intervalo  $I_1$  não foi estatisticamente significativo, além de ter apresentado elevado erro padrão em comparação às demais covariáveis, o que é coerente, uma vez que esse coeficiente não está associado a nenhum evento de interesse. Todas as demais covariáveis, com exceção de  $I_4$ , apresentaram coeficientes estatisticamente significativos, com valores  $-p$  inferiores a 5%.

As covariáveis APL, NNR e  $I_2$  apresentaram coeficientes negativos, indicando associação com a redução na chance de ocorrência do evento. Por outro lado, VIG e TEMP apresentaram coeficientes positivos e significativos, indicando associação com o aumento da chance de ocorrência do evento.

Podem ser obtidas as razões de chances de falha associadas a cada covariável, possibilitando a análise individual de cada coeficiente. Como exemplo, ao se aumentar em uma unidade a covariável APL mantendo-se as demais covariáveis constantes, obtém-se  $\exp(-0,0800) = 0,9231$ , ou seja, a taxa de produtividade passa a ser aproximadamente  $1 - 0,9231 = 0,0768$ , correspondendo a uma redução de cerca de 7,6% em relação à unidade anterior.

Para o coeficiente associado à covariável VIG, procede-se de forma semelhante: estima-se que o incremento de uma unidade nessa covariável, mantendo-se as demais constantes, resulte em um aumento de  $\exp(1,2217) = 3,3929$ , ou seja, aproximadamente 3,4 vezes a chance de produtividade da planta, indicando que plantas mais vigorosas tendem a apresentar maior produtividade.

Destaca-se que todas as covariáveis selecionadas possuem reconhecida importância para a produtividade dos cafezais ao longo do tempo. Pavan *et al.* (1994) ressaltam que a planta tende a reduzir a produtividade com o passar dos anos, em função da competição entre indivíduos, o que torna o coeficiente negativo coerente com os resultados obtidos.

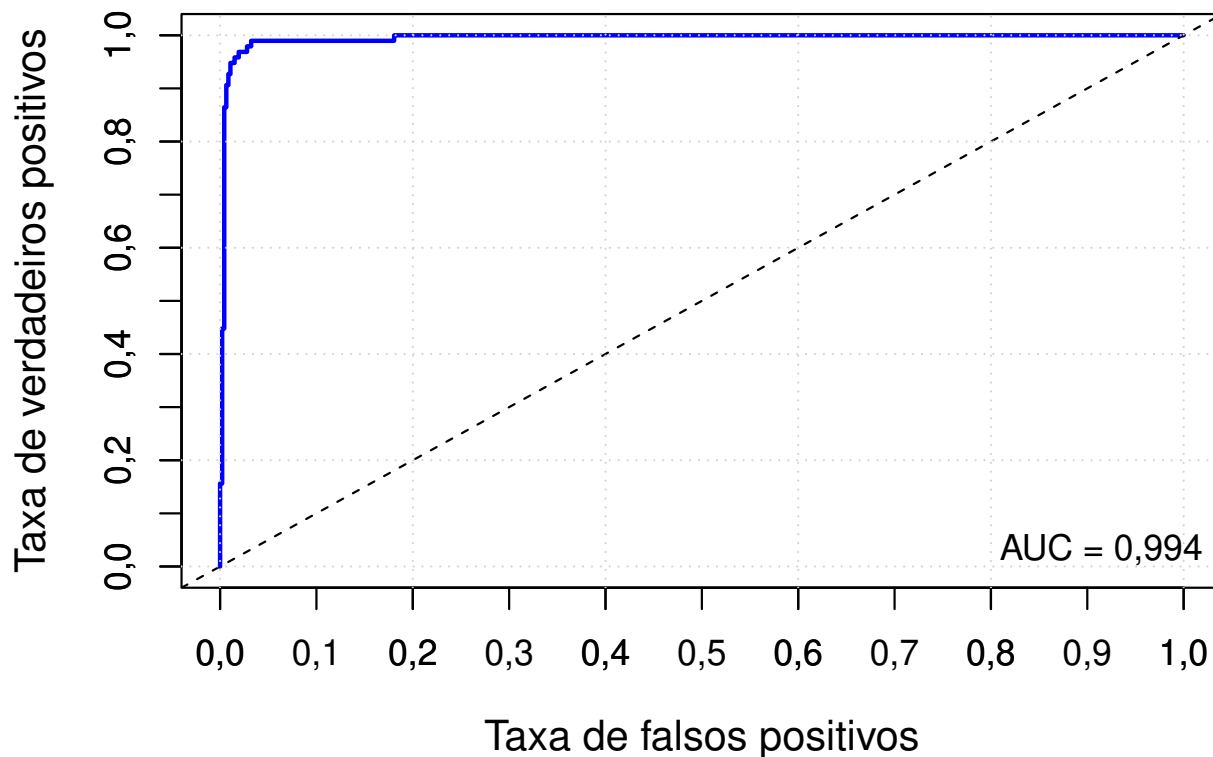
Adicionalmente, considerando o baixo investimento recebido pelo programa, é relevante mencionar que, quanto maior a planta, maior o investimento necessário para o incremento de produtividade.

Segundo Assad (2004), pequenas variações nos parâmetros climáticos podem influenciar significativamente a produtividade, o que reforça a importância da temperatura no processo produtivo da planta. A associação positiva entre temperatura e produtividade pode estar relacionada à reduzida quantidade de dados disponíveis.

O valor negativo associado à covariável NNR está diretamente relacionado à altura da planta: indivíduos mais velhos tendem a apresentar maior comprimento dos ramos plagiótropicos e, conseqüentemente, maior número desses ramos. Dessa forma, plantas maiores tendem a priorizar a alocação de nutrientes para funções vitais, e, na ausência de investimentos adequados, acabam não atingindo níveis elevados de produtividade. Miranda *et al.* (2005) destacam o papel dessas covariáveis no processo produtivo, corroborando os resultados obtidos neste estudo.

A curva ROC, apresentada na Figura (19), demonstra que o modelo possui elevada capacidade discriminatória, identificando com alta precisão as observações referentes às plantas que atingiram o evento de interesse, bem como aquelas que foram censuradas.

Figura 19 – Curva ROC, modelo logístico.



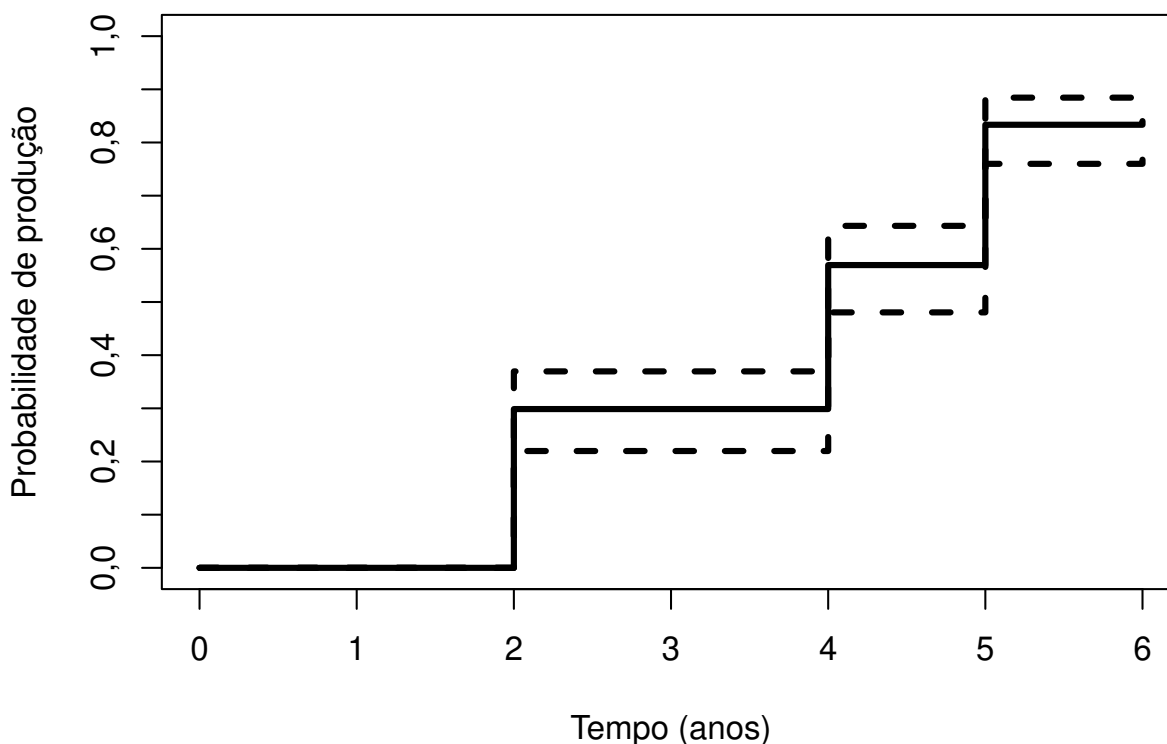
Fonte: Autor.

## 4.2 Catuaí Vermelho

Em comparação à testemunha Catuaí Vermelho IAC 144, observou-se que, dentre a população de 200 plantas analisadas, 121 (60,5%) apresentaram produtividade média igual ou superior à da testemunha. A análise de sobrevivência indicou um tempo mediano de ocorrência dos eventos igual a 2 anos, com maior frequência de eventos registrada em 2018.

A Figura 20 representa a curva de produtividade obtida pelo estimador de Kaplan-Meier juntamente com os respectivos intervalos de confiança.

Figura 20 – Produtividade estimada via Kaplan-Meier: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

A Tabela 27 apresenta um resumo detalhado dos tempos de produtividade das plantas, juntamente com os intervalos de confiança (*IC*) de 95%.

Tabela 27 – Produtividade estimada via Kaplan-Meier em comparação com a testemunha Catuaí Vermelho.

Tempo	n° risco	n° eventos	n° censura	Produtividade	Erro padrão	$IC_{95\%}$
0	200	0	56	0,00	0,00	(1,00; 1,00)
2	144	43	0	0,30	0,05	(0,78; 0,63)
4	112	39	0	0,67	0,09	(0,51; 0,35)
5	62	38	0	0,84	0,18	(0,24; 0,11)
6	24	1	23	0,85	0,19	(0,23; 0,10)

Como complemento à Figura 20, a Tabela 27 apresenta as observações censuradas nos tempos 0 e 6. As censuras no tempo 0 correspondem às plantas que morreram no início do

período de plantio, enquanto aquelas no tempo 6 referem-se às plantas que não atingiram produtividade média igual ou superior à da testemunha Catuaí Vermelho IAC 144 ao longo do acompanhamento.

No tempo 2 (ano de 2018), foram registradas 43 ocorrências do evento de interesse, correspondendo a uma probabilidade de 30% de ocorrência nesse intervalo. Nesse mesmo período, 112 plantas permaneciam sob risco, ou seja, passíveis de censura ou de alcançar produtividade média nos anos subsequentes. Esse período apresentou o maior número de ocorrências do evento no experimento. Em contraste, a testemunha Paraíso registrou seu pico de ocorrência no tempo 4 (ano de 2020).

No tempo 4 (ano de 2020), observaram-se 39 ocorrências do evento de interesse, resultando em uma probabilidade acumulada de 67% de ocorrência nesse período. Das 62 plantas remanescentes após esse intervalo, 38 apresentaram a ocorrência do evento no tempo 5 (ano de 2021), elevando a probabilidade acumulada para 84%. No tempo 6 (ano de 2022), foi registrada apenas uma nova ocorrência, enquanto que as 24 plantas restantes foram censuradas ao final do estudo.

A análise revelou que, do total de plantas monitoradas, 39,5% (79 indivíduos) foram censuradas ao longo do experimento, sendo 56 no tempo inicial (tempo 0) e 23 ao final do acompanhamento (tempo 6). Por outro lado, 60,5% da amostra (121 plantas) apresentaram produtividade média igual ou superior à da testemunha Catuaí Vermelho IAC 144, resultando em uma probabilidade acumulada de produção de 85% ao término do período experimental.

Observou-se que a produção das plantas manteve-se relativamente uniforme ao longo dos anos de acompanhamento. Os tempos 2 (2018), 4 (2020) e 5 (2021) registraram quantidades semelhantes de ocorrência do evento de interesse, com 43, 39 e 38 registros, respectivamente.

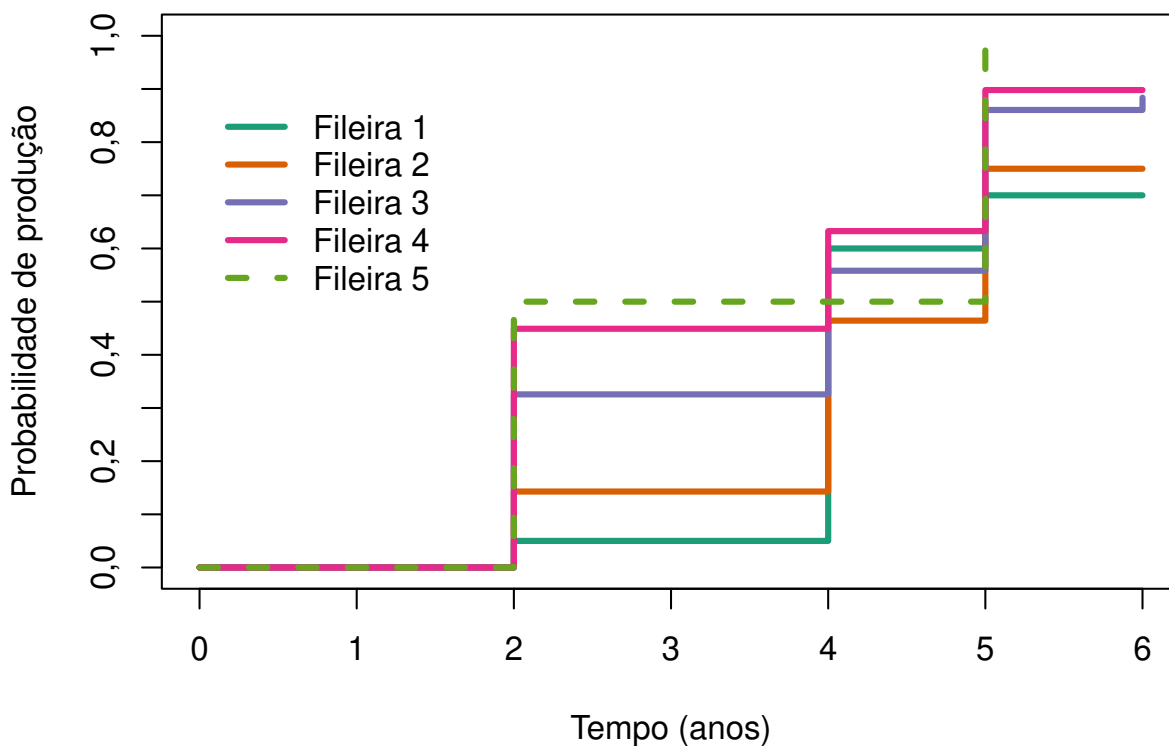
Os valores mais elevados de ocorrência do evento observados na comparação com a testemunha Catuaí Vermelho, em relação à testemunha Paraíso, refletem os parâmetros de referência utilizados para avaliação. A cultivar Catuaí Vermelho apresentou médias de produtividade inferiores ao longo do tempo em comparação à cultivar Paraíso, o que resultou em maior ocorrência do evento de interesse em sua comparação.

A probabilidade acumulada de 85% de produção ao longo do período de acompanhamento, evidencia o potencial do tratamento quanto à produtividade. Embora a testemunha Catuaí Vermelho tenha apresentado médias inferiores à variedade Paraíso ao longo do tempo, as condições de manejo adotadas durante o acompanhamento, aliadas à popularidade e ampla adoção do cultivar Catuaí Vermelho no território nacional, reforçam que os resultados obtidos foram promissores. Essa conclusão é fortalecida ao se considerar a totalidade de plantas que atingiram o evento de interesse.

#### 4.2.1 Comparação das curvas de produção

Observou-se que apenas a covariável fileira não rejeitou a hipótese nula de igualdade entre as curvas de produtividade, conforme indicado pelos testes de log-rank ( $\chi^2 = 7,6$ , valor  $-p = 0,1$ ) e Wilcoxon ( $\chi^2 = 8,8$ , valor  $-p = 0,07$ ). Esses resultados sugerem ausência de diferença estatisticamente significativa na probabilidade de produção entre as cinco fileiras analisadas. As curvas de produção correspondentes estão apresentadas na Figura 21.

Figura 21 – Produtividade estimada via Kaplan-Meier para cada nível da covariável fileira: comparação com a testemunha Catuaí Vermelho.

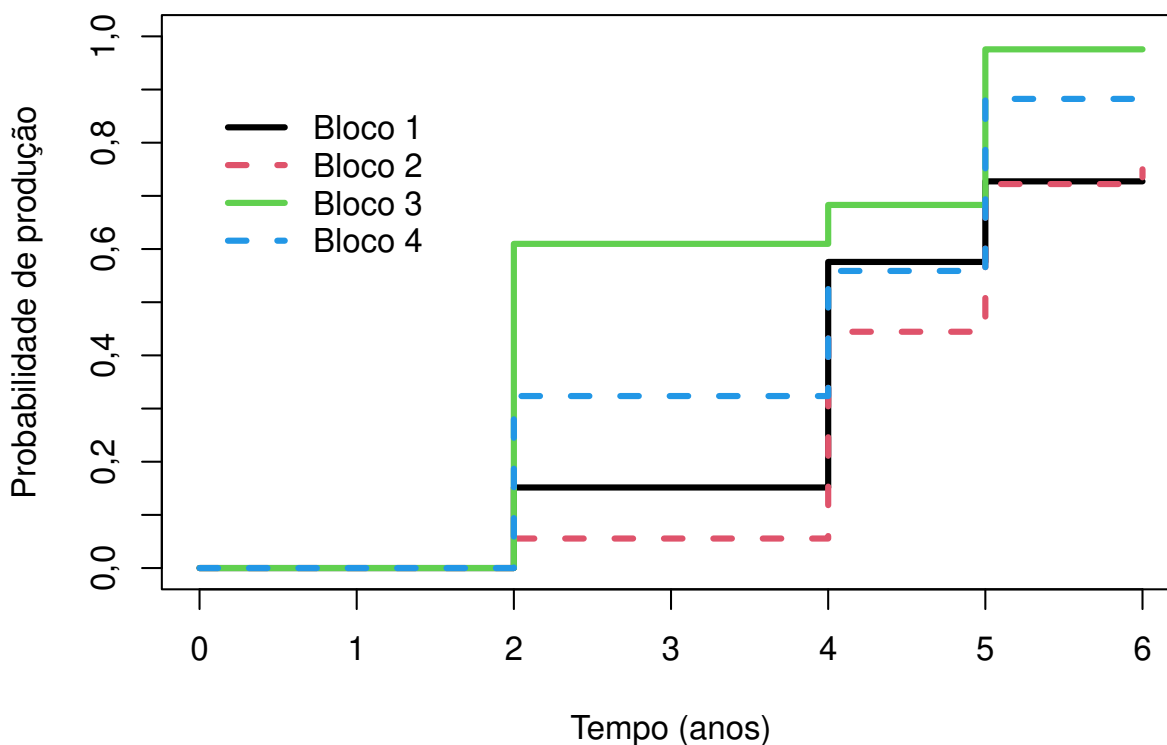


Fonte: Autor.

As demais covariáveis apresentaram diferenças significativas indicadas pelos dois testes, e a seguir serão apresentados os testes e as curvas de produção para cada nível das covariáveis.

A aplicação do teste de log-rank a covariável bloco evidenciou diferenças estatisticamente significativas entre as curvas de produtividade (valor- $p = 0,0003$ ), considerando um  $\alpha = 5\%$ . Da mesma forma o teste de Wilcoxon, mais sensível a diferenças iniciais, também detectou diferenças significativas (valor- $p < 0,0001$ ). Diante deste resultado, realizou-se uma análise comparativa dois a dois utilizando os testes de log-rank e Wilcoxon com o intuito de verificar quais níveis da covariável bloco diferem entre si, os resultados encontram-se detalhados na Tabela 28.

Figura 22 – Produtividade estimada via Kaplan-Meier para cada nível da covariável bloco: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Tabela 28 – Resultados dos testes de log-rank e Wilcoxon para comparação entre os níveis dos diferentes blocos.

Blocos	Teste de Log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Bloco 1 <i>versus</i> Bloco 2	0,2	0,6000	0,9	0,3000
Bloco 1 <i>versus</i> Bloco 3	9,5	<b>0,0020</b>	11,2	<b>0,0008</b>
Bloco 1 <i>versus</i> Bloco 4	1,7	0,2000	1,3	0,3000
Bloco 2 <i>versus</i> Bloco 3	14,8	<b>0,0001</b>	18,8	<b>&lt;0,0001</b>
Bloco 2 <i>versus</i> Bloco 4	3,7	0,0600	4,5	<b>0,0300</b>
Bloco 3 <i>versus</i> Bloco 4	3,4	0,0600	4,7	<b>0,0300</b>

Valores em negrito indicam que os grupos diferem entre si ( $p < 0,05$ ).

As comparações entre os blocos 1 e 3, bem como entre os blocos 2 e 3, apresentaram diferenças estatisticamente significativas nas curvas de produção ( $p < 0,05$ ) em ambos os testes aplicados. Essas diferenças podem ser atribuídas à posição dos blocos no campo experimental: enquanto os blocos 1 e 2 estão localizados nas extremidades, o bloco 3 encontra-se no centro

da lavoura, área caracterizada por maior concentração de matéria orgânica, o que favoreceu a superioridade produtiva das plantas ali cultivadas.

As comparações entre os blocos 2 e 4, e entre os blocos 3 e 4, apresentaram diferenças significativas apenas pelo teste de Wilcoxon. Conforme ilustrado na Figura 22, as curvas de produção exibem uma diferença acentuada no período inicial do acompanhamento, o que pode ter influenciado os resultados, uma vez que o teste de Wilcoxon atribui maior peso aos eventos ocorridos nas fases iniciais da análise.

As diferenças precoces identificadas justificam-se pela heterogeneidade no desenvolvimento das plantas durante as fases iniciais do experimento, diferença que foi reduzida ao final do acompanhamento, quando as plantas apresentaram produtividade mais homogênea.

Para a covariável ferrugem, o evento de interesse foi novamente registrado apenas nas categorias com notas 1 e 2, ambas associadas a plantas consideradas resistentes à doença. Esse resultado reforça a indicação de elevada resistência no material genético avaliado. No nível 1, foram identificadas 34 plantas, das quais 19 apresentaram produtividade média igual ou superior à da testemunha Catuaí Vermelho. No nível 2, foram classificadas 110 plantas, sendo que 82 delas atingiram o evento de interesse.

A ausência de ocorrência do evento de interesse nos níveis associados à manifestação da doença reforça a associação entre a produtividade e a redução dos danos causados pela ferrugem. As curvas de produção correspondentes a cada nível, bem como os resultados dos testes de comparação entre os níveis, encontram-se apresentados na Tabela 29 e na Figura 23.

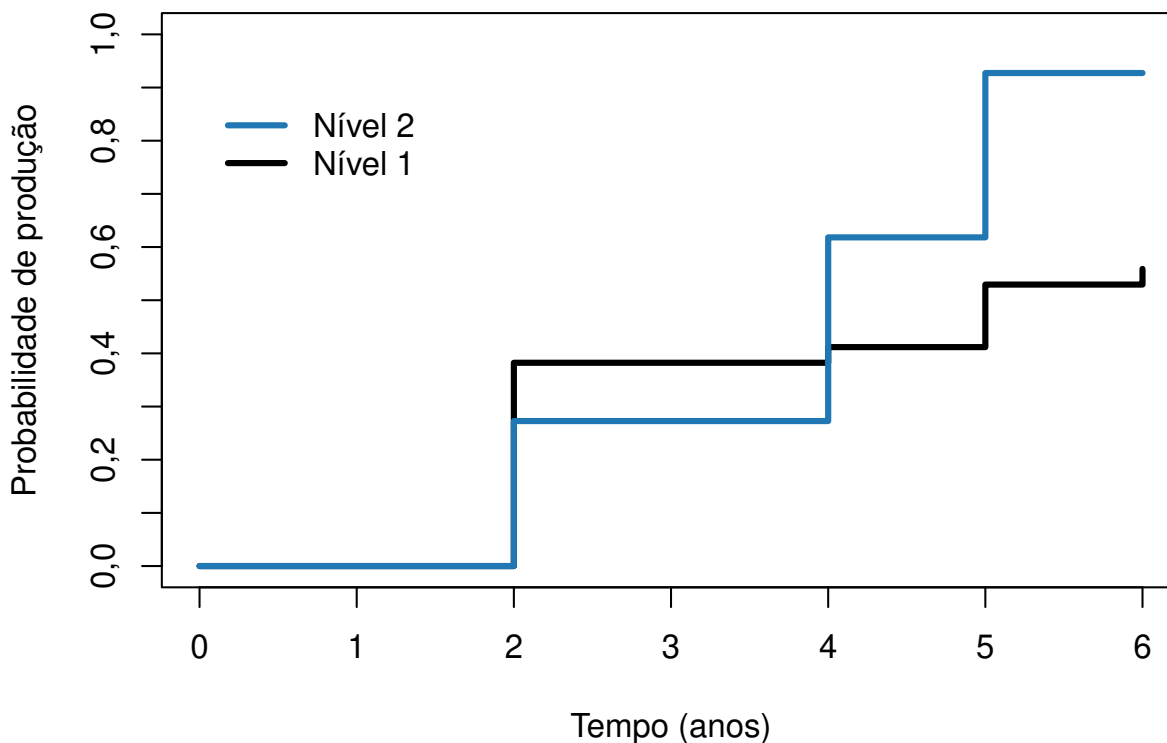
Tabela 29 – Testes de log-rank e Wilcoxon na comparação das curvas de produção de café sob diferentes níveis da covariável ferrugem.

Ferrugem	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Nível 1 <i>versus</i> Nível 2	38,8	<b>0,0007</b>	27,9	0,14

Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ).

Graficamente, observa-se que as curvas de produção dos dois níveis apresentam comportamento semelhante até o período correspondente a 4 anos, divergindo a partir desse ponto. Essa distinção no período final foi confirmada pelo teste de log-rank, que detectou diferenças estatisticamente significativas entre as curvas valor  $-p = 0,0007$ ) e  $\chi^2 = 38,8$ . Em contrapartida, o teste de Wilcoxon não identificou diferenças significativas valor  $-p = 0,14$ ) e  $\chi^2 = 27,9$  resultado consistente com sua maior sensibilidade para detectar diferenças iniciais entre as curvas de sobrevivência.

Figura 23 – Produtividade estimada via Kaplan-Meier para cada nível da covariável ferrugem: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Graficamente, observa-se que as curvas de produção dos dois níveis apresentam comportamento semelhante até o período correspondente a 4 anos, divergindo a partir desse ponto. Essa distinção no período final foi confirmada pelo teste de log-rank, que detectou diferenças estatisticamente significativas entre as curvas valor  $-p = 0,0007$ ) e  $\chi^2 = 38,8$ . Em contrapartida, o teste de Wilcoxon não identificou diferenças significativas valor  $-p = 0,14$ ) e  $\chi^2 = 27,9$  resultado consistente com sua maior sensibilidade para detectar diferenças iniciais entre as curvas de sobrevivência.

Esses resultados sugerem que a diferença entre os níveis analisados tornam-se mais evidentes após o quarto ano de observação, o qual plantas com nota 2 apresentam maior probabilidade de apresentarem a produtividade maior ou igual a testemunha Catuaí Vermelho.

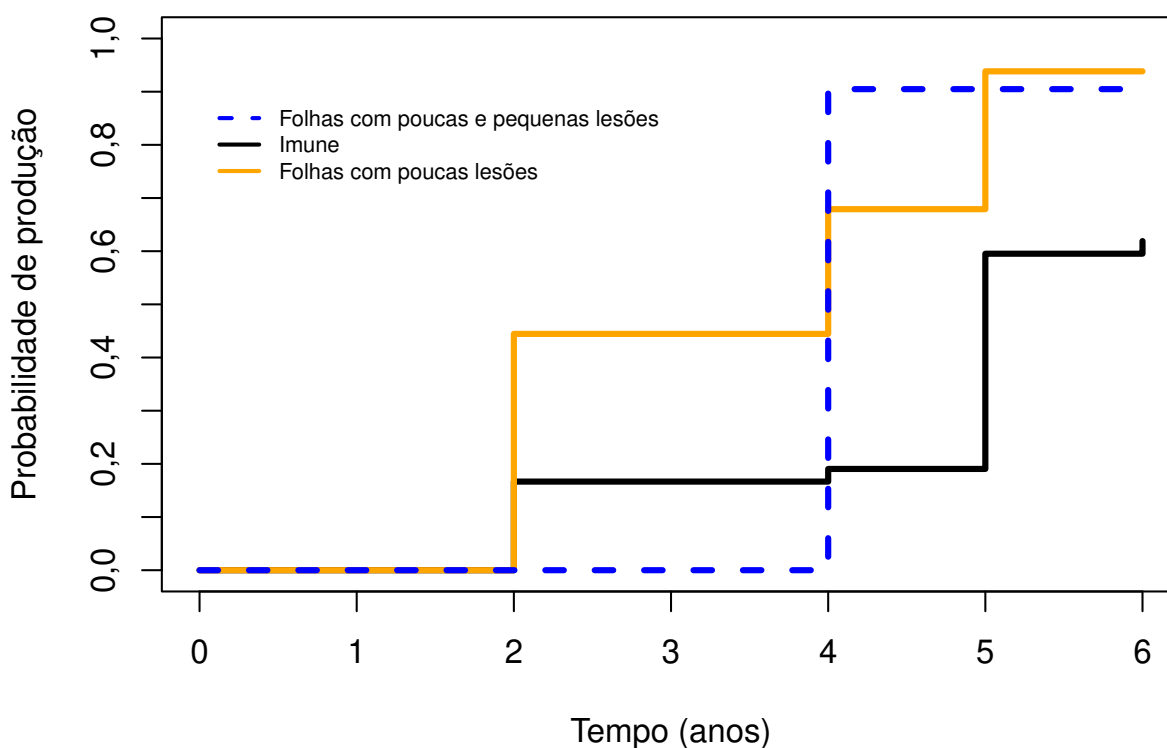
Para covariável Bicho mineiro, 56 plantas apresentaram nota 0, referente as plantas que morreram no início do estudo, 42 apresentaram notas 1 (plantas imunes a doença), sendo que 26 dessas plantas, alcançaram o evento de interesse. 81 plantas apresentaram nota 2 (folhas com poucas lesões), das quais 76 apresentaram o evento, ainda tiveram 21 plantas que apresentaram nota 3 (folhas com poucas e pequenas lesões), com 19 delas apresentando o evento de interesse.

Tabela 30 – Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável bicho-mineiro.

Bicho mineiro	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Nível 1 <i>versus</i> Nível 2	28,4	<b>&lt;0,0001</b>	25,3	<b>&lt;0,0001</b>
Nível 1 <i>versus</i> Nível 3	13,0	<b>&lt;0,0001</b>	13,4	<b>0,0003</b>
Nível 2 <i>versus</i> Nível 3	0,5	0,5000	2,1	0,1000

Valores em negrito indicam que os grupos diferem entre si ( $p < 0,05$ ). Nível 1: plantas imunes à doença; Nível 2: plantas com poucas lesões; Nível 3: plantas com poucas e pequenas lesões

Figura 24 – Produtividade estimada via Kaplan-Meier para cada nível da covariável bicho mineiro: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Ambos os testes, log-rank e Wilcoxon indicaram diferenças nas curvas de produção referentes as notas da covariável, os resultados referentes as comparações 2 a 2 e o valor da estatística dos testes de comparação aplicados na covariável BM são apresentados na Tabela 30 e na Figura 24.

Para a covariável cercosporiose (CER), 56 plantas apresentaram nota 0, correspondente àquelas que morreram no início do estudo. Nove plantas receberam nota 1 (sem presença de

cercosporiose), das quais seis atingiram o evento de interesse. No nível 2 (folhas com poucas lesões), 96 plantas foram classificadas, com 86 ocorrências do evento de interesse. No nível 3 (folhas com lesões médias), registraram-se 39 plantas, das quais 29 atingiram o evento.

O elevado número de plantas com notas indicativas de suscetibilidade à cercosporiose associa-se ao baixo nível de investimento empregado no experimento. Conforme destacado por Lima *et al.* (2018), a elevada incidência compromete a eficiência do manejo e reduz a capacidade de promover elevado vigor nas plantas.

Ambos os testes, log-rank e Wilcoxon, indicaram diferenças significativas entre as curvas de produção associadas aos níveis da covariável CER. Os resultados das comparações dois a dois, bem como os valores das estatísticas dos testes aplicados, encontram-se apresentados na Tabela 31.

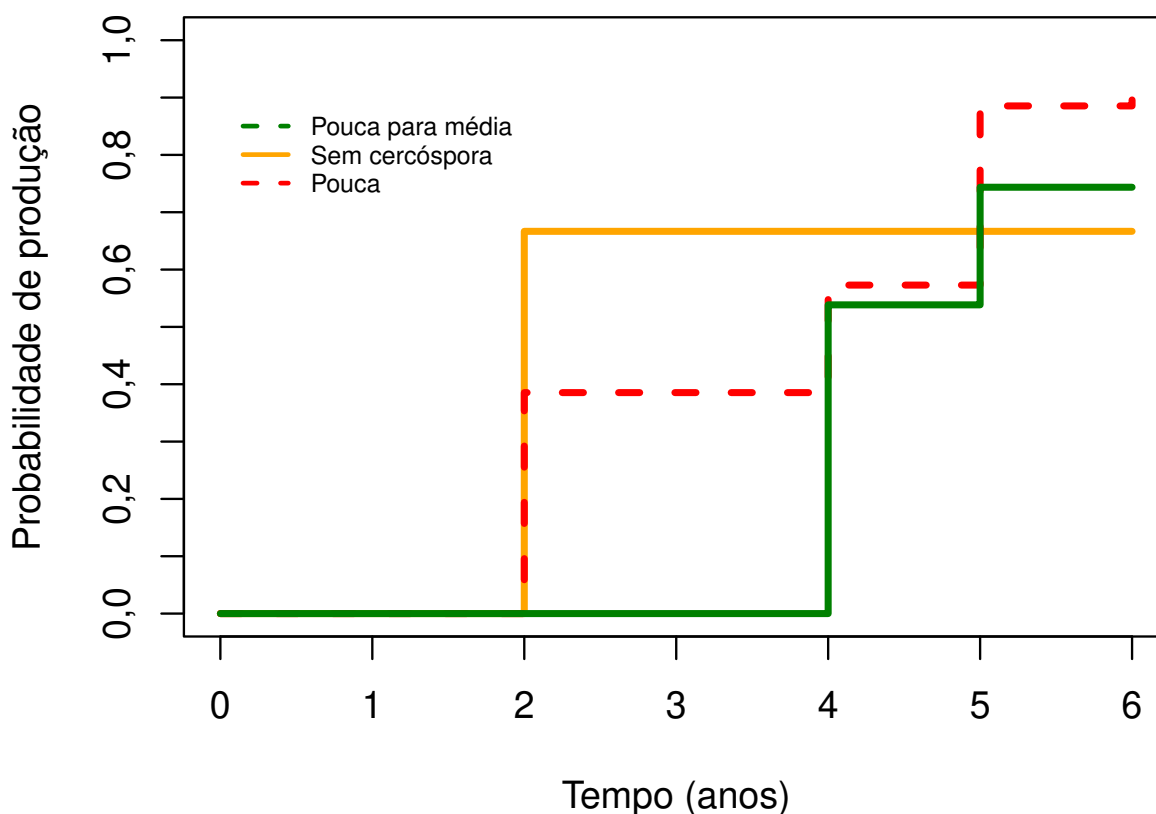
Tabela 31 – Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável cercosporiose.

Cercóspora	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Nível 1 <i>versus</i> Nível 2	0,2	0,700	0,4	0,500
Nível 1 <i>versus</i> Nível 3	2,0	0,200	6,1	<b>0,010</b>
Nível 2 <i>versus</i> Nível 3	5,0	<b>0,030</b>	6,8	<b>0,009</b>

Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ). Nível 1: plantas sem cercóspora; Nível 2: plantas com poucas lesões; Nível 3: plantas com poucas para médias lesões.

A Tabela 31 evidencia que não foram observadas diferenças significativas entre as curvas de produção associadas às notas 1 (sem cercosporiose) e 2 (folhas com poucas lesões) em nenhum dos testes aplicados. A Figura 25 ilustra um comportamento semelhante entre essas curvas ao longo do tempo, corroborando os resultados estatísticos. Para as curvas correspondentes às notas 1 e 3 (folhas com poucas a médias lesões), foram identificadas diferenças significativas pelo teste de Wilcoxon (valor  $-p < 0,05$ ), com destaque para uma diferença inicial acentuada, evidenciada graficamente e confirmada na Tabela 31. Em relação às curvas associadas às notas 2 e 3, ambos os testes indicaram diferenças significativas nas curvas de produtividade. Destaca-se que o nível 2 apresentou maior probabilidade de ocorrência de produtividade igual ou superior à da testemunha Catuaí Vermelho.

Figura 25 – Produtividade estimada via Kaplan-Meier para cada nível da covariável cercosporiose: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Os resultados mostraram que as plantas com pouca ou moderada presença de doenças conseguiram apresentar uma produtividade média igual ou superior à da testemunha Catuaí Vermelho. Diferentemente da comparação com a testemunha Paraíso, não foram encontradas plantas com notas 4 ou 5 para as covariáveis cercosporiose e bicho-mineiro, o que reforça a ideia de que, quanto menor a incidência de doenças, melhor o desempenho produtivo das plantas. Esses resultados vão ao encontro do que foi observado por Poma *et al.* (2024) e Avelino *et al.* (2015).

Assim como na comparação com a testemunha Paraíso, a covariável ferrugem demonstrou consistência em relação à resistência à doença. Apenas as notas 1 e 2 foram registradas, ambas associadas a plantas resistentes. Esses resultados indicam que a população avaliada é, em sua maioria, composta por indivíduos com resistência à ferrugem do cafeeiro. Tal evidência corrobora os achados de Mariz *et al.* (2025), que conseguiram combinar genes de resistência no tratamento e identificar a característica de resistência à ferrugem.

Além disso, mesmo plantas que apresentaram notas associadas a algum grau não imune às doenças cercosporiose e bicho-mineiro mostraram produtividade média igual ou superior à da testemunha Catuaí Vermelho, resultados promissores, considerando a ampla utilização do Catuaí Vermelho em todo o território nacional.

Para avaliação da qualidade da bebida, os testes estatísticos de log-rank e Wilcoxon rejeitaram a hipótese de igualdade entre as curvas de produção nos diferentes níveis da covariável uniformidade de maturação dos frutos. O teste log-rank apresentou estatística

$\chi^2 = 20,0$  (valor  $-p = 0,0002$ ), enquanto o teste Wilcoxon mostrou  $\chi^2 = 15,1$  (Valor  $-p = 0,002$ ).

Na análise, foram identificados cinco níveis da covariável. O nível 0 reuniu 56 plantas que não sobreviveram no início do plantio. No nível 1 (uniforme), foram registradas 5 plantas, nenhuma das quais apresentou o evento de interesse. O nível 2 (medianamente uniforme) incluiu 64 plantas, das quais 53 apresentaram o evento. Para o nível 3 (medianamente desuniforme), observaram-se 67 plantas, com 63 apresentando produtividade média igual ou superior à da testemunha Catuaí Vermelho. No nível 4 (desuniforme), contabilizaram-se 8 plantas, sendo que 5 apresentaram o evento de interesse.

Realizou-se uma comparação 2 a 2 entre os níveis, a fim de verificar quais deles apresentaram diferenças significativas. Os valores da estatística do teste  $\chi^2$  e os respectivos valores  $-p$  associados às comparações estão apresentados na Tabela 32. As curvas de produtividade correspondentes a cada nível encontram-se ilustradas na Figura 26.

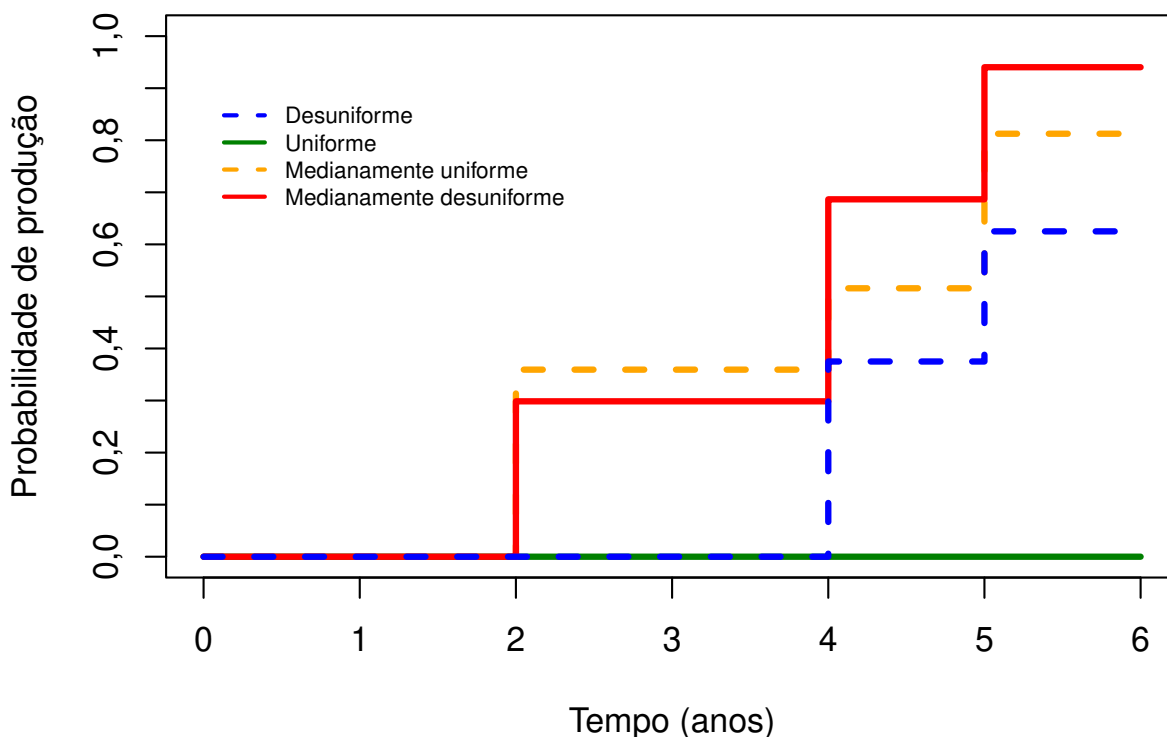
Tabela 32 – Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável uniformidade de maturação dos frutos.

Uniformidade de maturação dos frutos	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	$p$ -value	$\chi^2$	$p$ -value
Nota 1 <i>versus</i> Nota 2	10,3	<b>0,0010</b>	8,7	<b>0,0030</b>
Nota 1 <i>versus</i> Nota 3	16,8	<b>&lt;0,0001</b>	11,5	<b>0,0007</b>
Nota 1 <i>versus</i> Nota 4	4,5	<b>0,0300</b>	4,3	<b>0,0400</b>
Nota 2 <i>versus</i> Nota 3	3,1	0,0800	1,5	0,3000
Nota 2 <i>versus</i> Nota 4	1,9	0,2000	2,5	0,3000
Nota 3 <i>versus</i> Nota 4	6,3	<b>0,010</b>	5,3	<b>0,0200</b>

Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ). Nota 1: Uniforme Nota; Nota 2: Medianamente uniforme; Nota 3: Medianamente desuniforme; Nota 4: desuniforme.

Conforme indicado pela Tabela 32 as curvas associadas às plantas classificadas com nota 1 (uniforme) apresentaram diferenças estatisticamente significativas em relação às curvas dos níveis 2 (medianamente uniforme), 3 (medianamente desuniforme) e 4 (desuniforme), com valores  $-p$  inferiores a 0,005 em ambos os testes aplicados. A Figura 26 ilustra essas diferenças ao longo do tempo. Não foram observadas diferenças significativas entre as curvas dos níveis 2 e 3, bem como entre os níveis 2 e 4. Em contrapartida, as curvas dos níveis 3 e 4 apresentaram diferenças estatísticas em ambos os testes, sendo esse distanciamento também evidenciado graficamente na Figura 26, o que corrobora os resultados obtidos nas análises estatísticas.

Figura 26 – Produtividade estimada via Kaplan-Meier para cada nível da covariável uniformidade dos frutos: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

A diversidade observada na uniformidade de maturação dos frutos, junto com as diferenças entre as curvas nos testes, corrobora com o período de desuniformidade das características no cafeeiro. Melhores resultados e uma maturação mais uniforme, estão atrelados a um maior investimento, o que conseqüentemente torna o processo de seleção mais caro.

A análise das curvas correspondentes aos níveis da covariável tamanho de fruto (TF), com base nos testes de Log-Rank e Wilcoxon, resultou na rejeição da hipótese de igualdade entre as curvas de produção, com valores  $-p$  menores que 0,0001 e estatísticas  $\chi^2$  de 22,8 e 20,4, respectivamente.

Foram identificados quatro níveis da covariável tamanho do fruto. Um total de 56 plantas que não sobreviveram no início do estudo recebeu nota 0. Nenhuma planta foi classificada no nível 1 (miúdo). No nível 2 (pequeno), 22 plantas foram avaliadas, e 11 delas apresentaram o evento de interesse, correspondendo a 50%. No nível 3 (médio), 111 plantas foram incluídas, das quais 99 atingiram o evento (89,2%). Por fim, 11 plantas foram classificadas no nível 4 (grande), e todas apresentaram o evento de interesse, totalizando 100%.

A ausência de notas 5 se justifica, pois esse tratamento não possui tal característica em nenhum dos dois cultivares utilizados no cruzamento que o originou (Amarelo IAC 5162 e Híbrido de Timor MG0357). Por outro lado, a elevada frequência de notas 3 está associada aos parentais Híbridos de Timor, que apresentam essa característica.

Tabela 33 – Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável tamanho dos frutos.

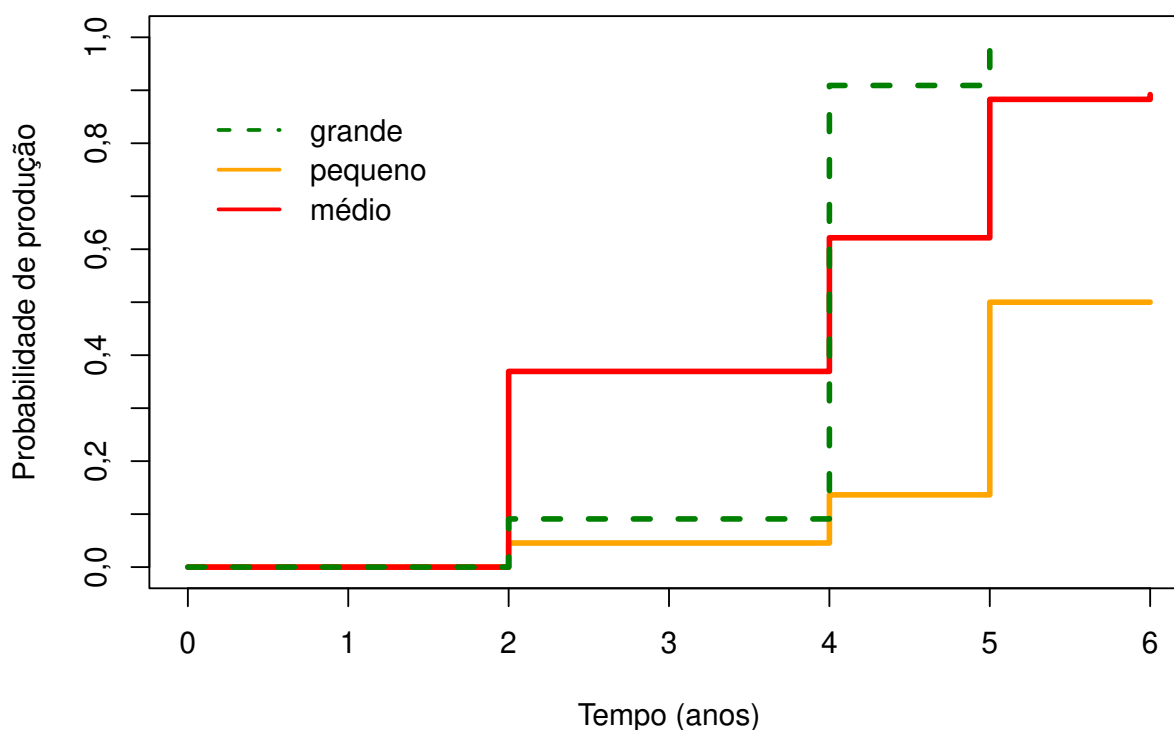
Tamanho do fruto	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
Nota 2 <i>versus</i> Nota 3	20,0	< <b>0,0001</b>	19,1	< <b>0,0001</b>
Nota 2 <i>versus</i> Nota 4	17,3	< <b>0,0001</b>	18,9	< <b>0,0001</b>
Nota 3 <i>versus</i> Nota 4	0,8	0,4000	0,0	0,9000

Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ). Nota 2: pequeno; Nota 3: médio; Nota 4: grande.

Para a verificação de diferenças entre os níveis da covariável tamanho do fruto, foram realizadas comparações 2 a 2 entre os níveis 2 (pequeno), 3 (médio) e 4 (grande). Os resultados obtidos para as estatísticas  $\chi^2$  e seus respectivos valores  $-p$  estão apresentados na Tabela 33, bem como as curvas de produção que são apresentadas na Figura 27.

As curvas de nível 3 (médio) e 4 (grande) não apresentaram diferenças em nenhum dos testes realizados, enquanto a curva de nível 2 (pequeno) diferiu das curvas 3 e 4, respectivamente como observado na Tabela 33.

Figura 27 – Produtividade estimada via Kaplan-Meier para cada nível da covariável tamanho do fruto: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Para covariável ciclo de maturação (CMT) os testes de Log-rank e Wilcoxon rejeitaram a hipótese de igualdade entre as curvas de produção entre os diferentes níveis da covariável. O

teste log-rank apresentou estatística  $\chi^2 = 23,5$  (valor  $-p = 0,0001$ ), enquanto o teste Wilcoxon mostrou  $\chi^2 = 19,4$  (valor  $-p = 0,0007$ ). No total, foram atribuídas seis notas à covariável em análise.

Verificou-se que 56 plantas receberam a nota 0, correspondente às plantas que morreram no início do estudo. A nota 1 associada a um ciclo de maturação (precoce) foi atribuída a 5 plantas, das quais uma delas apresentou produção média igual ou superior à testemunha Catuaí Vermelho. A nota foi observada em 39 plantas, sendo que 34 atingiram o evento de interesse. A nota 3 foi a mais frequente, atribuída a 87 plantas, com 79 delas apresentando o evento de interesse. Em relação à nota 4, 7 plantas foram classificadas nesse nível, com ocorrência do evento em 6 delas. Por fim, a nota 5 foi atribuída a 6 plantas, sendo registrada a ocorrência de apenas um evento de interesse para esse nível.

Tabela 34 – Testes de log-rank e Wilcoxon para comparação das curvas de produção de café em diferentes níveis da covariável ciclo de maturação.

Ciclo de maturação	Teste de log-rank		Teste de Wilcoxon	
	$\chi^2$	$p$ -value	$\chi^2$	$p$ -value
Nota 1 <i>versus</i> Nota 2	8,4	<b>0,004</b>	7,6	<b>0,0060</b>
Nota 1 <i>versus</i> Nota 3	11,6	<b>&lt;0,0001</b>	9,1	<b>0,0030</b>
Nota 1 <i>versus</i> Nota 4	5,3	<b>0,0200</b>	5,2	<b>0,0100</b>
Nota 1 <i>versus</i> Nota 5	0,0	0,9000	0,0	0,9000
Nota 2 <i>versus</i> Nota 3	0,1	0,8000	0,1	0,8000
Nota 2 <i>versus</i> Nota 4	0,1	0,8000	0,5	0,5000
Nota 2 <i>versus</i> Nota 5	10,4	<b>0,0010</b>	9,3	<b>0,0020</b>
Nota 3 <i>versus</i> Nota 4	0,4	0,5000	0,5	0,5000
Nota 3 <i>versus</i> Nota 5	14,0	<b>&lt;0,0002</b>	11,1	<b>0,0008</b>
Nota 4 <i>versus</i> Nota 5	8,9	<b>0,0030</b>	6,4	<b>0,0100</b>

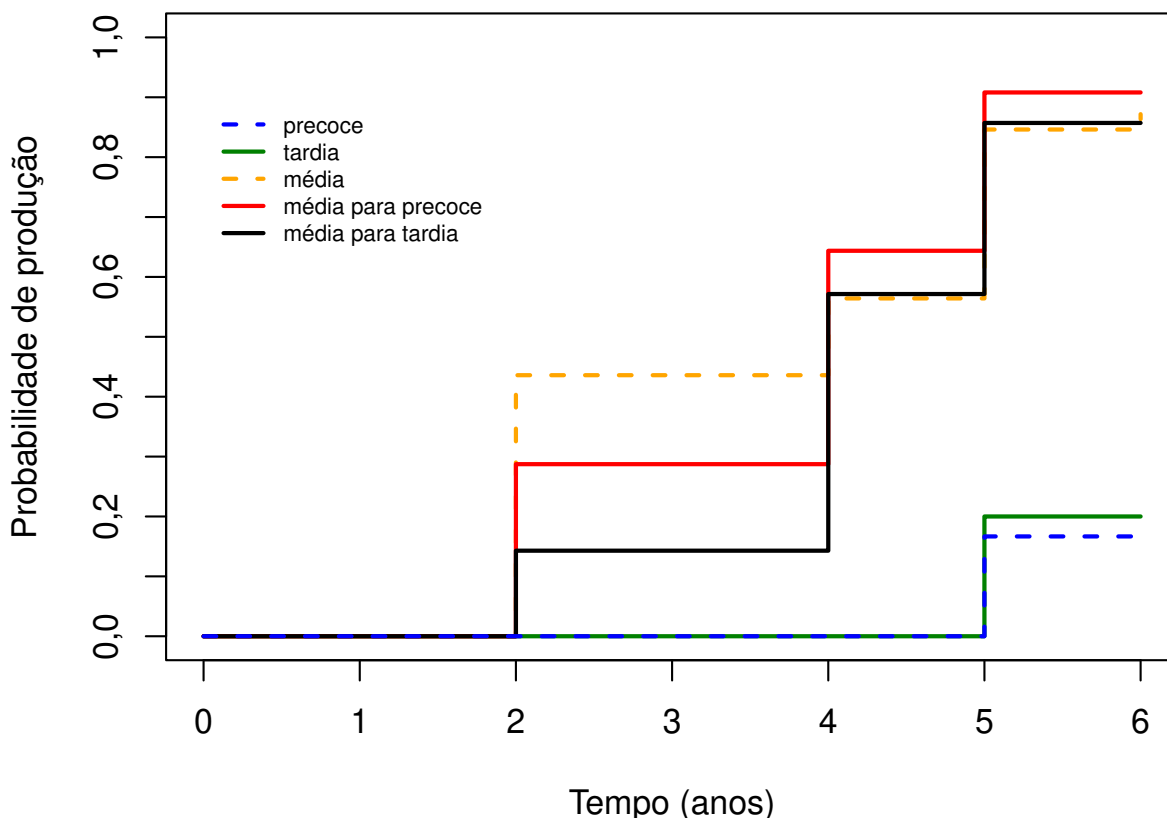
Valores em negrito indicam que os grupos diferem entre si (valor  $-p < 0,05$ ). Nota 1: precoce; Nota 2: média para precoce; Nota 3: média; Nota 4: média para tardia; Nota 5: tardia.

Foram realizados testes 2 a 2 para os níveis de 1 a 5 da covariável ciclo de maturação, na Tabela 34 encontram-se os testes  $\chi^2$  e seus respectivos valores  $-p$ , na Figura 28 tem-se as respectivas curvas de produção, associadas a cada nível da covariável.

As curvas associadas às notas 2 (média para precoce), 3 (média) e 4 (média para tardia) não apresentaram diferenças significativas em nenhum dos testes aplicados, o mesmo ocorrido ao comparar as curvas correspondentes às notas 1 (precoce) e 5 (tardia). Conforme apresentado na Tabela 34, as plantas classificadas com nota 1 não diferiram daquelas com notas 3 e 4, enquanto as plantas com nota 5 diferiram de todos os demais níveis, exceto do nível 1.

A Figura 28 apresenta as curvas de produtividade correspondentes aos diferentes níveis da covariável ciclo de maturação (CMT). Observa-se um comportamento semelhante ao longo do tempo entre os níveis 2, 3 e 4, bem como entre os níveis 1 e 5, o que é corroborado pelos resultados dos testes estatísticos descritos na Tabela 34.

Figura 28 – Produtividade estimada via Kaplan-Meier para cada nível da covariável ciclo de maturação: comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Os diferentes ciclos de maturação identificados mostram-se promissores. Os níveis precoce e tardio apresentaram curvas de produção mais baixas, mas com comportamentos semelhantes ao longo do tempo, conforme apontado pelos testes estatísticos. Já as curvas associadas aos ciclos médio, médio para tardio e médio para precoce não apresentaram diferenças significativas na probabilidade de produção ao longo do tempo.

Esse padrão observado permite que os produtores escalonem a colheita em diferentes fases, otimizando o uso da mão de obra e melhorando a qualidade da bebida, uma vez que favorece a colheita de frutos mais uniformes. Além disso, como as curvas de produção mantêm uma probabilidade constante ao longo do tempo, o escalonamento possibilita colheitas com altos volumes em cada etapa, garantindo eficiência e qualidade ao processo produtivo.

#### 4.2.2 Modelo final utilizando aproximação de Breslow

O modelo selecionado incluiu as covariáveis Altura da planta (APL), Vigor (VIG), Temperatura (TEMP) e precipitação (PREC). Todas as etapas do processo de seleção de covariáveis encontram-se disponíveis na Tabela 35.

Tabela 35 – Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Breslow.

Passo	Modelo	Loglike	TRV	Valor- <i>p</i>
1	1	NULO	1107,4017	–
2		APL	1073,8861	33,5156
3		BL	1103,9527	3,4490
4		BM	1095,5191	11,8826
5		CER	1104,8825	2,5192
6		CRP	1107,2535	0,1482
7		DCA	1005,1839	102,2177
8		DCO	1107,3741	0,0275
9		FER	1100,1425	7,2591
10		FIL	1103,2433	4,1584
11		NNR	1107,2288	0,1729
12		NRP	1107,2288	0,3094
13		VIG	1076,1282	31,2735
14		TEMP	976,2884	131,1132
15		PREC	1000,6232	106,7784
16	2	APL+BL+BM+DCA+FER+FIL+NNR+VIG+TEMP+PREC	932,5672	–
17		BL+BM+DCA+FER+FIL+NNR+VIG+TEMP+PREC	940,5197	7,9526
18		APL+BM+DCA+FER+FIL+NNR+VIG+TEMP+PREC	933,7395	1,1723
19		APL+BL+DCA+FER+FIL+NNR+VIG+TEMP+PREC	933,4628	0,8956
20		APL+BL+BM+FER+FIL+NNR+VIG+TEMP+PREC	932,6222	0,0551
21		APL+BL+BM+DCA+FIL+NNR+VIG+TEMP+PREC	933,1986	0,6315
22		APL+BL+BM+DCA+FER+NNR+VIG+TEMP+PREC	934,3605	1,7934
23		APL+BL+BM+DCA+FER+FIL+VIG+TEMP+PREC	934,1472	1,5800
24		APL+BL+BM+DCA+FER+FIL+NNR+TEMP+PREC	937,9794	5,4123
25		APL+BL+BM+DCA+FER+FIL+NNR+VIG+PREC	961,4188	28,8517
26		APL+BL+BM+DCA+FER+FIL+NNR+VIG+TEMP	939,0936	6,5265
27	3	APL+VIG+PREC+TEMP	939,6611	–
28		APL+VIG+PREC+TEMP+BL	939,3864	0,2747
29		APL+VIG+PREC+TEMP+BM	937,8406	1,8205
30		APL+VIG+PREC+TEMP+DCA	939,6126	0,0485
31		APL+VIG+PREC+TEMP+FER	938,5464	1,1147
32		APL+VIG+PREC+TEMP+FIL	938,0735	1,5876
33		APL+VIG+PREC+TEMP+NNR	936,5942	3,0669
34	4	APL+VIG+PREC+TEMP	939,6611	–
35		APL+VIG+PREC+TEMP+CER	937,2949	2,3662
36		APL+VIG+PREC+TEMP+CRP	939,6220	0,0391
37		APL+VIG+PREC+TEMP+DCO	939,3409	0,3202
38		APL+VIG+PREC+TEMP+NNR	934,3605	3,0669
39		APL+VIG+PREC+TEMP+NRP	937,5380	2,1231
40	5	APL+VIG+PREC+TEMP	939,6611	–
41		VIG+PREC+TEMP	946,0192	6,3581
42		APL+PREC+TEMP	999,4894	59,8283
43		APL+VIG+PREC	969,8844	30,2233
44		APL+VIG+TEMP	947,8394	8,1783
	Modelo Final	APL+VIG+TEMP+PREC	936,6611	–

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Para avaliar a contribuição individual de cada covariável no modelo, construiu-se a Tabela 36, baseada no Teste da Razão de Verossimilhanças (TRV). A tabela apresenta a estatística qui-quadrado calculada ( $\chi^2_{\text{cal}}$ ), os respectivos graus de liberdade (gl) e os valores  $-p$  associados a cada teste.

Todas as covariáveis selecionadas mostraram-se estatisticamente significativas valor  $-p < 0,05$ , com melhorias consistentes nos valores da log-verossimilhança, indicando seu poder explicativo no modelo. Um resumo dos coeficientes e respectivos intervalos de confiança de

Tabela 36 – Teste da razão de verossimilhança para o modelo de aproximações de Breslow: Comparação com a testemunha Catuaí Vermelho.

	Log-verossimilhança	$\chi^2_{\text{cal}}$	gl	Valor- $p$
NULO	-553,70			
APL	-536,94	33,515	1	< <b>0,0001</b>
VIG	-514,88	44,125	1	< <b>0,0001</b>
TEMP	-473,92	81,928	1	< <b>0,0001</b>
PREC	-469,83	8,178	1	<b>0,0042</b>

Valores em negrito indicam que a contribuição da inclusão das covariáveis no modelo é significativa (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

95% encontram-se sumarizados na Tabela 37.

Tabela 37 – Análise dos coeficientes, razão de risco e intervalos de confiança para o modelo Breslow: Comparação com a testemunha Catuaí Vermelho.

Variável	Coefficiente	Razão de riscos	Erro padrão	Valor- $p$	$IC_{95\%}$
APL	-0,0186	0,9816	0,0042	< <b>0,0001</b>	(0,9736; 0,9896)
VIG	0,3124	1,3667	0,1248	<b>0,01230</b>	(1,0702; 1,7455)
TEMP	2,2575	9,5600	1,1415	<b>0,0479</b>	(1,0204; 89,5666)
PREC	-0,0059	0,9940	0,0020	<b>0,0037</b>	(0,9900; 0,9981)

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Wald (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Podemos destacar que o modelo de riscos proporcionais de Cox indicou que a covariável vigor das plantas (VIG) apresentou efeito positivo e significativo sobre a probabilidade da produtividade ser igual ou superior à testemunha Catuaí Vermelho. Especificamente, o aumento no vigor das plantas elevou em aproximadamente 94% a chance de alcançar produtividade média igual ou superior à testemunha. Por outro lado, as covariáveis altura da planta (APL) e precipitação (PREC) apresentaram efeitos negativos significativos. O aumento de APL reduziu em cerca de 1,7% a chance de produtividade ser igual ou superior à testemunha, enquanto o aumento da precipitação reduziu a chance de ocorrência do evento em aproximadamente 5,7%. A temperatura apresentou o maior valor para razão de riscos, indicando que uma variação de uma unidade aumenta em 9,6 vezes a ocorrência do evento.

### 4.2.3 Adequação do modelo

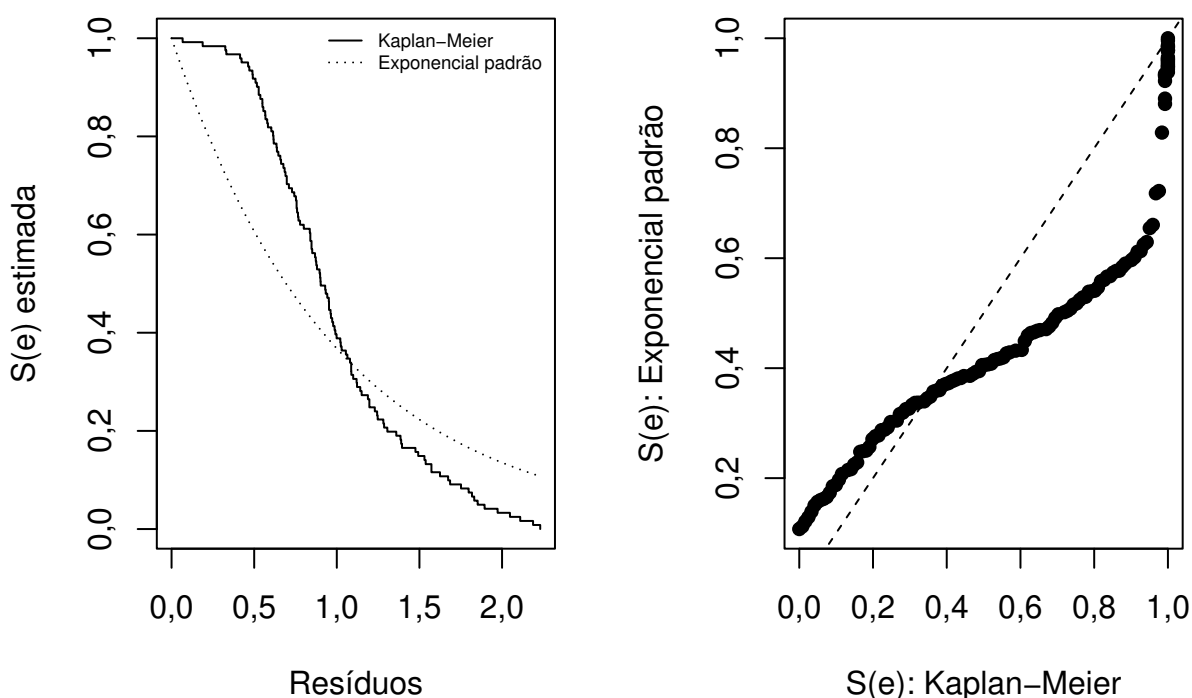
A Tabela 38 apresenta os resultados do teste de Schoenfeld aplicado às covariáveis do modelo. Verificou-se que a covariável VIG atende às pressuposições de riscos proporcionais. No entanto, as demais covariáveis, APL e DCA, assim como o ajuste global do modelo, apresentaram violações dessa pressuposição, evidenciadas pelos (valores  $-p < 0,05$ ), indicando quebra da suposição de riscos proporcionais.

Tabela 38 – Teste de Schoenfeld aplicado às covariáveis do modelo de aproximação de Breslow: Comparação com a testemunha Catuaí Vermelho.

	$\chi^2_{\text{cal}}$	gl	Valor- $p$
APL	18,10	1	<b>&lt;0,0001</b>
VIG	1,57	1	0,7356
TEMP	7,65	1	<b>0,0002</b>
PREC	7,65	1	0,1457
GLOBAL	21,26	3	<b>0,0001</b>

Valores em negrito indicam que as covariáveis violam a suposição de riscos proporcionais, pois rejeitam a hipótese nula do teste de Schoenfeld (valor  $-p < 0,05$ ). APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação; GLOBAL: Ajuste geral para o modelo completo.

Figura 29 – Análise dos resíduos de Cox-Snell do modelo de aproximação Breslow: Comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

A análise de resíduos dos gráficos de Cox-Snell apresentada na Figura 29 mostra que eles não seguem distribuição exponencial padrão, o que também evidencia a não adequação do modelo de Cox com aproximação Breslow aos dados do estudo.

#### 4.2.4 Modelo final utilizando aproximação de Efron

O processo de seleção de covariáveis resultou na escolha das mesmas covariáveis identificadas para o modelo com aproximação de Breslow: altura da planta (APL), vigor (VIG),

temperatura (TEMP) e precipitação (PREC). As etapas referentes ao processo de seleção encontram-se descritas na Tabela 39.

Tabela 39 – Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação de Efron: Comparação com a testemunha Catuaí Vermelho.

Passo	Modelo	Loglike	TRV	Valor- <i>p</i>	
1	1	NULO	1046.9017	-	-
2		APL	996.7502	50.1515	<0,0001
3		BL	1040.8750	6.0267	0.0141
4		BM	1027.8950	19.0067	<0,0001
5		CER	1043.7053	3.1964	0.0738
6		CRP	1046.9017	0.0000	0.9962
7		DCA	887.7986	159.1031	<0,0001
8		DCO	1046.9017	0.0000	0.9993
9		FER	1031.9907	14.9110	<0,0001
10		FIL	1039.9495	6.9522	0.0084
11		NNR	1046.8723	0.0294	0.8638
12		NRP	1046.8723	0.4132	0.5204
13		VIG	999.4157	47.4861	<0,0001
14		TEMPER	823.5749	223.3268	<0,0001
15		PREC	881.0680	165.8337	<0,0001
16	2	APL+BL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	744.6447	-	-
17		BL+BM+CER+CRP+DCA+FER+FIL+VIG+TEMPER+PREC	771.7967	27.1520	<0,0001
18		APL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	748.4266	3.7818	0.0518
19		APL+BL+CER+CRP+DCA+FER+FIL+VIG+TEMPER+PREC	745.7918	1.1471	0.2842
20		APL+BL+BM+DCA+FER+FIL+VIG+TEMPER+PREC	746.5829	1.9382	0.1639
21		APL+BL+BM+CER+FER+FIL+VIG+TEMPER+PREC	744.8344	0.1897	0.6632
22		APL+BL+BM+CER+DCA+FIL+VIG+TEMPER+PREC	745.5056	0.8608	0.3535
23		APL+BL+BM+CER+DCA+FER+VIG+TEMPER+PREC	749.8515	5.2067	0.0225
24		APL+BL+BM+CER+DCA+FER+FIL+TEMPER+PREC	757.0157	12.3709	<b>0.0021</b>
25		APL+BL+BM+CER+DCA+FER+FIL+VIG+PREC	806.4126	61.7679	<0,0001
26		APL+BL+BM+CER+DCA+FER+FIL+VIG+TEMPER	761.9863	17.3416	<0,0001
27	3	APL+VIG+TEMPER+PREC+FIL	753.8513	-	-
28		APL+VIG+TEMPER+PREC+FIL+BL	750.5961	3.2552	0.0712
29		APL+VIG+TEMPER+PREC+FIL+BM	750.7651	3.0862	0.0790
30		APL+VIG+TEMPER+PREC+FIL+CER	751.0990	2.7523	0.0971
31		APL+VIG+TEMPER+PREC+FIL+DCA	753.8147	0.0366	0.8482
32		APL+VIG+TEMPER+PREC+FIL+FER	752.3259	1.5254	0.2168
33	4	APL+VIG+TEMPER+PREC+FIL	753.8513	-	-
34		APL+DCA+VIG+TEMPER+CRP	753.8413	0.0101	0.9200
35		APL+VIG+TEMPER+PREC+FIL+DCO	752.9184	0.9330	0.3341
36		APL+VIG+TEMPER+PREC+FIL+NRP	748.0650	5.7863	0.0162
37		APL+VIG+TEMPER+PREC+FIL+NNR	749.3570	4.4944	0.0340
38	5	APL+VIG+TEMPER+PREC+FIL+NRP+NNR	745.7682	-	-
39		VIG+TEMPER+PREC+FIL+NRP+NNR	760.6519	14.8837	<0,0001
40		APL+TEMPER+PREC+FIL+NRP+NNR	757.8341	12.0659	<0,0001
41		APL+VIG+PREC+FIL+NRP+NNR	801.9641	56.1959	<0,0001
42		APL+VIG+TEMPER+FIL+NRP+NNR	801.9641	13.3643	<0,0001
43		APL+VIG+TEMPER+PREC+NRP+NNR	749.4760	3.7078	0.0542
44		APL+VIG+TEMPER+PREC+FIL+NNR	749.3570	3.5888	0.0582
45		APL+VIG+TEMPER+PREC+FIL+NRP	748.0650	2.2968	0.1296
50	Modelo Final	APL+VIG+TEMP+PREC	757,3301	-	-

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Tabela 40 – Teste da razão de verossimilhança para o modelo de aproximações de Efron: Comparação com a testemunha Catuaí Vermelho.

	Log-verossimilhança	$\chi_{\text{cal}}^2$	gl	Valor- $p$
NULO	-523,45			
APL	-498,38	50,515	1	< <b>0,0001</b>
VIG	-461,27	74,205	1	< <b>0,0001</b>
TEMP	-385,64	151,272	1	< <b>0,0001</b>
PREC	-378,67	13,944	1	<b>0,0001</b>

Valores em negrito indicam que a contribuição da inclusão da covariável no modelo é significativa (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

A Tabela 40 apresenta a avaliação da contribuição individual de cada covariável ao modelo nulo, com base no Teste da Razão de Verossimilhanças (TRV). São apresentados a estatística qui-quadrado ( $\chi_{\text{cal}}^2$ ), os respectivos graus de liberdade (gl) e os valores- $p$  associados a cada teste.

Todas as covariáveis selecionadas mostraram evidências estatísticas de melhora no ajuste do modelo (valor  $-p < 0,05$ ), indicando contribuições relevantes para sua explicação. Esses resultados também se refletem em melhorias consistentes na log-verossimilhança. A Tabela 41 apresenta os coeficientes estimados para cada covariável e as razões de risco acompanhados de seus respectivos intervalos de confiança de 95%.

Tabela 41 – Análise dos coeficientes, razão de risco e intervalos de confiança para o modelo Efron: Comparação com a testemunha Catuaí Vermelho.

Variável	Coeficiente	Razão de riscos	Erro padrão	Valor- $p$	$IC_{95\%}$
APL	-0,0265	0,9737	0,0042	< <b>0,0001</b>	(0,9657; 0,9820)
VIG	0,4627	1,5884	0,1269	<b>0,0003</b>	(1,2386; 2,0371)
TEMP	2,7442	15,5523	0,9713	<b>0,0047</b>	(2,3174; 104,3750)
PREC	-0,0070	0,9929	0,0017	< <b>0,0001</b>	(0,9895; 0,9964)

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Wald (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

#### 4.2.5 Adequação do modelo

A Tabela 42 apresenta os resultados do teste de Schoenfeld aplicado às covariáveis do modelo. Verificou-se que a covariável VIG atende às pressuposições de riscos proporcionais. No entanto, as demais covariáveis, APL, TEMP e PREC, assim como o ajuste global do modelo, apresentaram violações dessa pressuposição, evidenciadas pelos (valores  $-p < 0,05$ ), indicando quebra da suposição de riscos proporcionais.

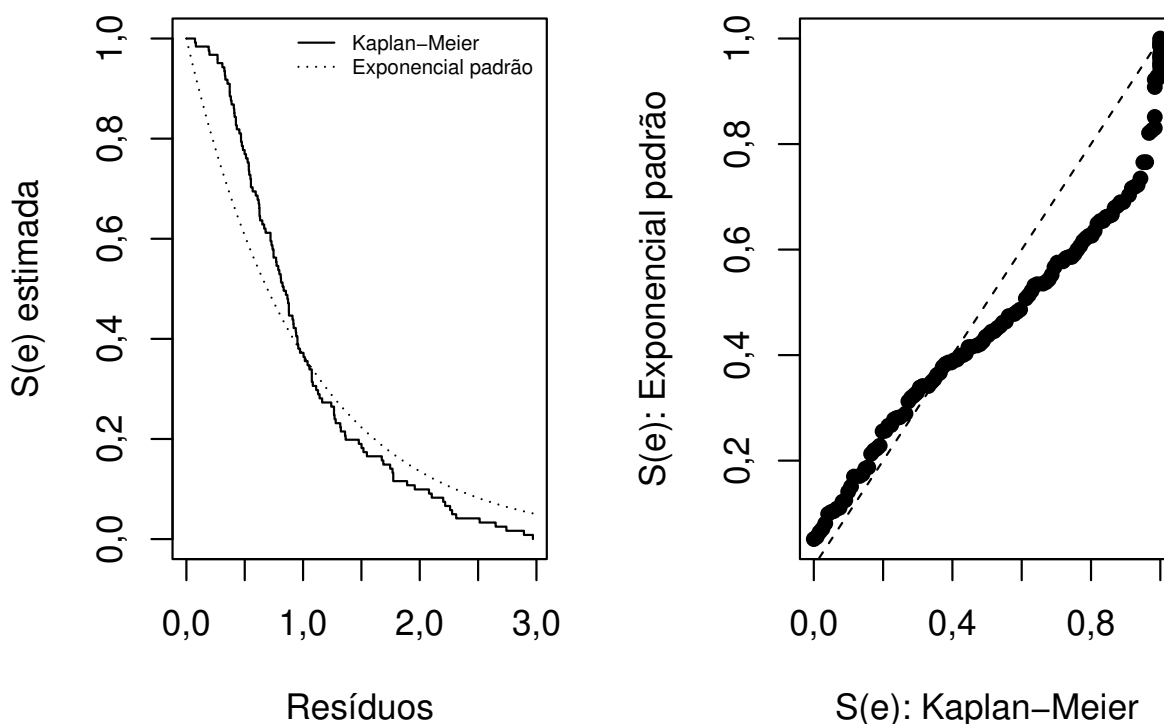
Tabela 42 – Teste de Schoenfeld aplicado as covariáveis do modelo de aproximação de Efron: Comparação com a testemunha Catuaí Vermelho.

	$\chi^2_{\text{cal}}$	gl	Valor- $p$
APL	34,3195	1	< <b>0,0001</b>
VIG	0,0428	1	0,836
TEMP	43,0416	1	< <b>0,0001</b>
PREC	7,8909	1	< 0,005
GLOBAL	56,9598	4	< <b>0,0001</b>

Valores em negrito indicam que as covariáveis não atendem as pressuposições de riscos proporcionais (valor  $p < 0,05$ ). APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação; GLOBAL: Ajuste incluindo todas as covariáveis.

Mesmo apresentando um ajuste superior em relação ao modelo que utilizou aproximações Breslow, a análise de resíduos dos gráficos de Cox-Snell apresentada na Figura 30 mostra que eles não seguem distribuição exponencial padrão, o que também evidencia a não adequação do modelo de Cox aos dados do estudo.

Figura 30 – Análise dos resíduos de Cox-Snell do modelo de aproximação Efron: Comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

#### 4.2.6 Modelo final utilizando a aproximação exata

Ao final do processo de seleção de covariáveis, optou-se por permanecer com o modelo

referente a linha 27 da Tabela 43. O modelo final selecionado incluiu as covariáveis Altura da planta (APL), Temperatura (TEMP), precipitação (PREC). Todas as etapas do processo de seleção de covariáveis encontram-se disponíveis na Tabela 43.

Tabela 43 – Seleção de covariáveis utilizando o modelo de regressão de Cox, com aproximação exata: Comparação com a testemunha Catuaí Vermelho.

PASSO	MODELOS	Looglike	TRV	P_VALOR
1	1 NULO	384,6356	-	-
2	APL	325,7504	58,8852	< <b>0,0001</b>
3	BL	366,1264	18,5092	0,0003
4	BM	351,9998	32,6358	< <b>0,0001</b>
5	CER	378,8819	5,7538	0,0563
6	CRP	384,3722	0,2635	0,6077
7	DCA	215,0767	169,5590	< <b>0,0001</b>
8	DCO	384,5863	0,0494	0,8241
9	FER	372,5293	12,1064	0,0005
10	FIL	376,8899	7,7457	0,1013
11	NNR	384,3321	0,3035	0,5817
12	NRP	384,3321	0,5488	0,4588
13	VIG	329,1160	55,5197	< <b>0,0001</b>
14	TEMPER	161,5737	223,0620	< <b>0,0001</b>
15	PREC	217,6633	166,9724	< <b>0,0001</b>
16	2 APL+BL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	20,1898	-	-
17	BL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	49,6593	29,4695	< <b>0,0001</b>
18	APL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	23,5001	3,3103	0,1010
19	APL+BL+CER+DCA+FER+FIL+VIG+TEMPER+PREC	18,2127	1,9771	0,5772
20	APL+BL+BM+DCA+FER+FIL+VIG+TEMPER+PREC	23,5697	3,3799	0,3367
21	APL+BL+BM+CER+FER+FIL+VIG+TEMPER+PREC	20,2190	0,0293	0,8641
22	APL+BL+BM+CER+DCA+FER+FIL+VIG+TEMPER+PREC	21,0677	0,8779	0,6447
23	APL+BL+BM+CER+DCA+FER+VIG+TEMPER+PREC	27,9439	7,7542	0,1010
24	APL+BL+BM+CER+DCA+FER+FIL+TEMPER+PREC	22,7255	2,5357	0,1113
25	APL+BL+BM+CER+DCA+FER+FIL+VIG+PREC	83,0665	62,8768	< <b>0,0001</b>
26	APL+BL+BM+CER+DCA+FER+FIL+VIG+TEMPER	27,8645	7,6747	0,0056
27	3 APL+TEMP+PREC	51,6727	-	-
28	APL+TEMP+PREC+BLO	46,0742	5,5985	0,1329
29	APL+TEMP+PREC+BM	42,5975	9,0752	0,0283
30	APL+TEMP+PREC+CER	34,5027	17,1700	0,0007
31	APL+TEMP+PREC+DCA	51,6668	0,0059	0,9388
32	APL+TEMP+PREC+FER	46,7795	4,8932	0,0270
33	APL+TEMP+PREC+FIL	39,7191	11,9536	0,0270
34	APL+TEMP+PREC+VIG	43,7074	7,9653	0,0048
35	4 APL+VIG+PREC+TEMPER+BM+CER+FER+FIL	23,7332	-	-
36	APL+VIG+PREC+TEMPER+BM+CER+FER+FIL+CRP	23,2809	0,4523	0,5013
37	APL+VIG+PREC+TEMPER+BM+CER+FER+FIL+DCO	22,9881	0,7451	0,3880
38	APL+VIG+PREC+TEMPER+BM+CER+FER+FIL+NNR	21,7082	2,0250	0,1547
39	APL+VIG+PREC+TEMPER+BM+CER+FER+FIL+NRP	22,0895	1,6437	0,1998
40	5 APL+VIG+PREC+TEMPER+BM+CER+FER+FIL	23,7332	-	-
41	VIG+PREC+TEMPER+BM+CER+FER+FIL	65,6180	41,8848	< <b>0,0001</b>
42	APL+PREC+TEMPER+BM+CER+FER+FIL	26,3296	2,5964	0,1071
43	APL+VIG+PREC+BM+CER+FER+FIL	84,9851	61,2519	< <b>0,0001</b>
44	APL+VIG+TEMPER+BM+CER+FER+FIL	55,8547	32,1214	< <b>0,0001</b>
45	APL+VIG+PREC+TEMPER+CER+FER+FIL	22,8309	0,9023	0,8249
46	APL+VIG+PREC+TEMPER+BM+FER+FIL	26,6119	2,8787	0,4107
47	APL+VIG+PREC+TEMPER+BM+CER+FIL	24,4271	0,6939	0,7068
48	APL+VIG+PREC+TEMPER+BM+CER+FER	28,2467	4,5135	0,3410
49	Modelo Final APL+TEMP+PREC	51,6727	-	-

Valores em negrito indicam a inclusão ou exclusão de covariáveis significativas para o modelo (valor  $-p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; BL: Blocos; BM: Bicho mineiro; CER: Cercosporiose; CRP: Comprimento dos ramos plagiotrópicos; DCA: Diâmetro do caule; DCO: Diâmetro da copa; FER: Ferrugem; FIL: Fileiras; NNR: Número de nós nos ramos plagiotrópicos; NRP: Número de ramos plagiotrópicos; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação.

Para avaliar a contribuição individual de cada covariável no modelo, construiu-se a Tabela 44, baseada no Teste da Razão de Verossimilhanças (TRV). A tabela apresenta a estatística qui-quadrado calculada ( $\chi^2_{\text{cal}}$ ), os respectivos graus de liberdade (gl) e os valores  $-p$  associados

a cada teste.

Tabela 44 – Teste da razão de verossimilhança para o modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho.

	Log-verossimilhança	$\chi_{cal}^2$	gl	Valor- $p$
NULO	-553,70			
APL	-162,875	58,885	1	< <b>0,0001</b>
TEMP	-25,836	114,345	1	< <b>0,0001</b>
PREC	-83,009	159,733	1	< <b>0,0001</b>

Valores em negrito indicam que a contribuição da inclusão das covariáveis no modelo é significativa (valor- $p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; TEMP: Temperatura; PREC: Precipitação.

Todas as covariáveis selecionadas mostraram-se estatisticamente significativas valor- $p < 0,05$ , com melhorias consistentes nos valores da log-verossimilhança, indicando seu poder explicativo no modelo. Um resumo dos coeficientes e respectivos intervalos de confiança de 95% encontram-se sumarizados na Tabela 45.

Tabela 45 – Análise dos coeficientes, razão de risco e respectivos intervalos de confiança para o modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho.

Variável	Coefficiente	Razão de riscos	Erro padrão	Valor- $p$	$IC_{95\%}$
APL	-0,1434	0,8664	0,0294	< <b>0,0001</b>	(0,8179; 0,9179)
TEMP	9,5857	14,540	1,1415	< <b>0,0001</b>	(145,5073; 1453000)
PREC	-0,0059	0,9753	0,0054	< <b>0,0001</b>	(0,9649; 0,9858)

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Wald (valor- $p < 0,05$ ). NULO: Modelo nulo; APL: Altura da planta; TEMP: Temperatura; PREC: Precipitação.

O aumento de APL reduziu em cerca de 13,36% o risco da produtividade ser igual ou superior à testemunha, enquanto o aumento da precipitação reduziu a chance de ocorrência do evento em aproximadamente 2,47%. A temperatura apresentou o maior valor para razão de riscos, indicando que uma variação de uma unidade aumenta em 14,5 vezes a ocorrência do evento.

#### 4.2.7 Adequação do modelo

A Tabela 46 apresenta os resultados do teste de Schoenfeld aplicado às covariáveis do modelo. Verificou-se que as covariáveis, APL, PREC e TEMP assim como o ajuste global do modelo, apresentaram violação da pressuposição de riscos proporcionais, evidenciadas pelos (valores- $p < 0,05$ ).

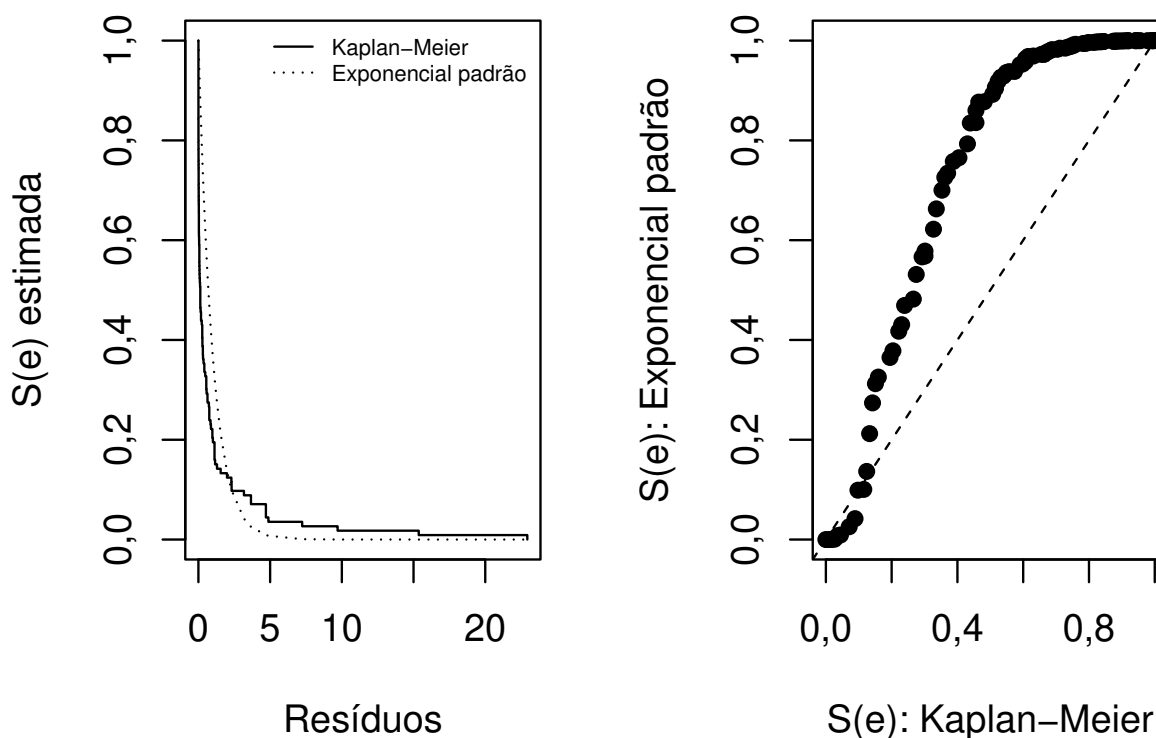
Tabela 46 – Teste de Schoenfeld aplicado às covariáveis do modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho.

	$\chi^2_{cal}$	gl	Valor- $p$
APL	18,10	1	<b>&lt;0,0001</b>
TEMP	7,65	1	<b>&lt;0,0001</b>
PREC	7,65	1	<b>&lt;0,0001</b>
GLOBAL	21,26	3	<b>0,0001</b>

Valores em negrito indicam que as covariáveis violam a suposição de riscos proporcionais, pois rejeitam a hipótese nula do teste de Schoenfeld (valor  $-p < 0,05$ ). APL: Altura da planta; VIG: Vigor; TEMP: Temperatura; PREC: Precipitação; GLOBAL: Ajuste geral para o modelo completo.

A análise de resíduos dos gráficos de Cox-Snell apresentada na Figura 31 mostra que eles não seguem distribuição exponencial padrão, o que também evidencia a não adequação do modelo de Cox aos dados do estudo.

Figura 31 – Análise dos resíduos de Cox-Snell do modelo de aproximação exata: Comparação com a testemunha Catuaí Vermelho.



Fonte: Autor.

Apesar da metodologia de seleção ter identificado covariáveis relevantes para os modelos, as aproximações adotadas não satisfizeram os pressupostos de riscos proporcionais exigidos pelo modelo de Cox, tampouco apresentaram bom ajuste com base na análise dos resíduos de Cox-Snell. A elevada proporção de empates observada nos dados contribuiu para essas

limitações, indicando a necessidade de adoção de modelos discretos como alternativa mais apropriada.

#### 4.2.8 Modelos finais

Inicialmente, elaborou-se a Tabela 47, que descreve os tempos de produção das plantas analisadas. Percebe-se que três, dos cinco intervalos analisados apresentaram tempos de produção empatados, com um total de 59% das plantas, o que corrobora a utilização de modelos discretos.

Tabela 47 – Resumo dos dados apresentados pela tabela de vida: Comparação com a testemunha Catuaí Vermelho.

Intervalo de tempo	Total sob risco	Número de eventos	Número de censuras	$\hat{S}(t)$	Erro padrão de $\hat{S}(t)$
(0, 2]	200	0	56	1,00	–
(2, 4]	144	43	0	0,70	0,054
(4, 5]	101	39	0	0,43	0,095
(5, 6]	62	38	0	0,16	0,186
[6, +∞)	24	1	23	0,15	0,191

Após a realização do processo de seleção *stepwise*, o modelo final selecionado incluiu as covariáveis Fileira (FIL), Vigor vegetativo (VIG), Altura da planta (APL), Temperatura (TEMP), Número de nós nos ramos plagiotrópicos (NNR), Precipitação (PREC) e os intervalos de tempo  $I_1$  [2016,2018),  $I_2$  [2018,2020),  $I_4$  [2021,2022) e  $I_5$  [2022, +∞), como indicado na Tabela 48. Observou-se que  $I_1$  apresentou erro padrão relativamente alto, o que é coerente, dado que esse coeficiente não apresentou nenhum evento de interesse.

Optou-se pela utilização do modelo de Cox, em função dos menores valores de AIC (59,274) e BIC (119,121), em comparação aos valores obtidos com o modelo logístico (AIC = 62,807; BIC = 122,655). Sob o modelo de Cox, observa-se que a taxa de produtividade da  $FIL_3$  foi aproximadamente  $\exp(1,3180) = 3,735$  vezes maior do que a da fileira  $FIL_1$ , com interpretação análoga para as demais fileiras.

Verificou-se que os coeficientes associados às covariáveis VIG, APL, TEMP, PREC, NNR, bem como aos intervalos de tempo  $I_2$ ,  $I_4$ ,  $I_5$  e à fileira  $FIL_2$ , apresentaram-se estatisticamente significativos pelo teste z, considerando um nível de significância de  $\alpha = 5\%$ . As demais covariáveis não apresentaram significância estatística.

Tabela 48 – Estimativas dos parâmetros dos modelos finais do processo *stepwise* para os modelos de Cox e logístico: Comparação com a testemunha Catuaí Vermelho.

	Modelo de Cox			Modelo logístico		
	Estimativa	Erro padrão	Valor $-p$	Estimativa	Erro padrão	Valor $-p$
$FIL_2$	2,9950	1,4200	<b>0,0350</b>	4,8900	1,9800	<b>0,0135</b>
$FIL_3$	1,3180	0,9615	0,1706	2,1020	1,4420	0,1451
$FIL_4$	1,6340	1,0660	0,1252	2,6780	1,4420	0,0935
$FIL_5$	3,2520	3,7300	0,3832	6,0500	6,4400	0,3476
$I_1$	-33,5800	2415,0000	0,9889	-39,4700	2321,0000	0,9864
$I_2$	-10,3900	2,6080	<b>&lt;0,0001</b>	-14,3700	3,8300	<b>0,0002</b>
$I_4$	15,2300	4,1910	<b>0,0003</b>	21,0700	6,7720	<b>0,0019</b>
$I_5$	18,0800	6,1470	<b>0,0033</b>	25,5000	8,7280	<b>0,0035</b>
VIG	1,4610	0,5604	<b>0,0091</b>	1,8780	0,7204	<b>0,0091</b>
APL	-0,1023	0,0296	<b>0,0005</b>	-0,1416	0,0410	<b>0,0006</b>
TEMP	8,4230	2,7280	<b>0,0020</b>	11,3700	3,8670	<b>0,0033</b>
NNR	-0,1991	0,0961	<b>0,0383</b>	-0,2486	0,1343	0,0642
PREC	-0,0189	0,0054	<b>0,0005</b>	-0,0264	0,07328	<b>0,0008</b>

Valores em negrito indicam que os coeficientes diferem estatisticamente de zero pelo teste Z (valor  $-p < 0,05$ ).

$FIL_2$ : Fileira de número 2;  $FIL_3$ : Fileira de número 3;  $FIL_4$ : Fileira de número 4;  $FIL_5$ : Fileira de número 5;  $I_1$ : Intervalo de tempo referente ao período de [2016, 2018];  $I_2$ : Intervalo de tempo referente ao período de [2018, 2020];  $I_4$ : Intervalo de tempo referente ao período de [2021, 2022];  $I_5$ : Intervalo de tempo referente ao período de [2022,  $+\infty$ ); VIG: Vigor vegetativo das plantas; APL: Altura da planta; TEMP: Temperatura; NNR: Número de nós nos ramos plagiotrópicos; PREC: Precipitação.

O coeficiente estimado para a variável NNR foi negativo, indicando que ao considerar as demais covariáveis fixas, aumentos em NNR estão associados a uma redução na taxa de risco de produtividade. A razão de riscos associada ao incremento unitário em NNR foi calculada como  $\exp(-0,1991) = 0,819$ , ou seja, o risco de produtividade passa a ser aproximadamente 0,819 vezes comparado ao risco anterior. Consequentemente, a redução percentual é dada por:  $10 - 8,19 = 1,81 = 18,1\%$ .

Embora esse resultado possa, inicialmente, parecer contraditório, a estimativa negativa observada para APL corrobora esse padrão. Tal comportamento pode ser atribuído ao baixo investimento no programa de melhoramento, que induz as plantas a priorizarem a alocação de nutrientes para funções fisiológicas essenciais, em detrimento da produtividade. Esse contexto justifica os sinais negativos observados nos coeficientes de NNR e APL. Além disso, a baixa produtividade nos tempos finais do intervalo amostrado contribuiu para esses achados.

Segundo Assad (2004), pequenas variações nos parâmetros climáticos podem influenciar significativamente a produtividade, o que corrobora a importância da temperatura e da precipitação no processo produtivo das plantas. O elevado valor do coeficiente associado à estimativa da variável TEMP e sua associação positiva com a resposta podem estar relacionados à reduzida quantidade de dados disponíveis, bem como à unidade de medida adotada (temperatura média anual).

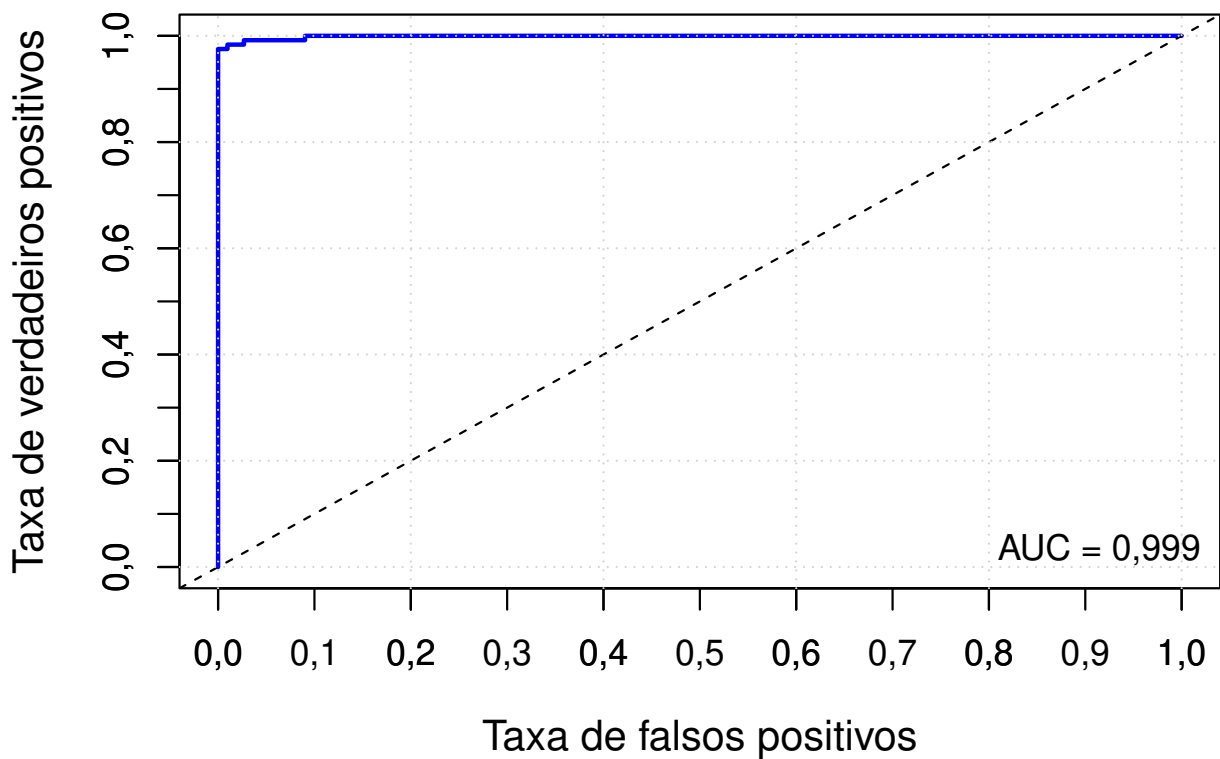
Embora tenha apresentado coeficiente negativo, a precipitação (PREC) demonstrou-se relevante no processo produtivo. A variação, ainda que ínfima, indica que o incremento de uma unidade na covariável, mantendo-se as demais constantes, resulta em uma redução

aproximada de 1,9% na chance de produtividade em relação ao valor anterior.

No que se refere à covariável Vigor (VIG), o aumento de uma unidade, considerando-se as demais constantes, representa um incremento de aproximadamente 4,3 vezes na chance de produtividade em relação à unidade anterior, o que reforça a associação entre plantas mais vigorosas e uma produtividade superior em comparação às demais.

A curva ROC representada na Figura 32 indica que o modelo consegue capturar com precisão as observações referentes as plantas que atingiram o evento de interesse e as que foram censuradas.

Figura 32 – Curva ROC, modelo de riscos proporcionais.



Fonte: Autor.

## 5 CONCLUSÕES

A análise de sobrevivência indicou que o cruzamento do Híbrido de Timor MG0357 com o Tupi Amarelo IAC 5162 apresentou um alto potencial produtivo em comparação às testemunhas Paraíso e Catuaí Vermelho.

Inicialmente apresentado por Mariz *et al.* (2025), o presente estudo indicou que o cultivar se mostrou resistente à ferrugem do cafeeiro, além de apresentar uma boa capacidade produtiva, mesmo em plantas que manifestaram algum nível de ocorrência das doenças cercosporiose e bicho-mineiro.

Os testes de log-rank e Wilcoxon se mostraram pertinentes, identificando diferenças tardias e precoces nas distintas curvas de produção das covariáveis categóricas analisadas. Destaca-se, em especial, a covariável ciclo de maturação, cujas curvas associadas aos ciclos média, média para precoce e média para tardia não apresentaram diferenças significativas em termos de probabilidade de produção, sendo, contudo, mais produtivas que aquelas associadas aos ciclos precoce e tardio.

Os modelos de Cox que utilizam aproximações da função de verossimilhança parcial para lidar com empates não atenderam às pressuposições de riscos proporcionais, além de não apresentarem um bom ajuste, sendo, portanto, não recomendados para este tipo de análise.

Por outro lado, os modelos discretos de riscos proporcionais, bem como o modelo logístico, apresentaram bom ajuste e elevada capacidade discriminatória entre plantas que atingiram o evento e aquelas censuradas, mostrando-se ideais para este tipo de situação. As covariáveis altura da planta, diâmetro do caule, temperatura, precipitação, número de nós nos ramos plagiotrópicos e vigor vegetativo destacaram-se como as mais importantes para explicar a produtividade do tratamento nos modelos selecionados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716-723, 1974.
- ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DO CAFÉ - ABIC. **Indicadores da Indústria de Café | 2023**. ABIC. 2023. Disponível em: <https://estatisticas.abic.com.br/estatisticas/indicadores-da-industria/indicadores-da-industria-de-cafe-2023>. Acesso em: 20 de maio. 2024.
- ARAQUE-SALAZAR, H.; DUQUE, H. Variables agronómicas determinantes de la productividad del cultivo de café en fincas del departamento de Caldas. **Revista Cenicafé**, [S. l.], v. 70, n. 1, p. 81–92, 2019. DOI: 10.38141/10778/70106. Disponível em: <https://publicaciones.cenicafe.org/index.php/cenicafe/article/view/64>. Acesso em: 28 abr. 2025.
- ASSAD, E. D.; PINTO, H. S.; ZULLO JUNIOR, J.; ÁVILA, A. M. H. Impacto das mudanças climáticas no zoneamento agroclimático do café no Brasil. **Pesquisa agropecuária brasileira**, v. 39, p. 1057-1064, 2004.
- AVELINO, J.; CRISTANCHO, M.; GEORGIU, S.; IMBACH, P.; AGUILAR, L.; BORNEMANN, G.; LÄDERACH, P.; ANZUETO, F.; HRUSKA, A. J.; MORALES, C. The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. **Food security**, v. 7, p. 303-321, 2015.
- BOTELHO, C. E.; NADALETI, D. H. S.; DOS SANTOS SILVA, J.; CARVALHO, G. R.; PEREIRA, A. A.; DE OLIVEIRA, A. C. B.; FERREIRA, A. D. **Cultivar de café MGS Paraíso 2: potencial de produtividade e qualidade na região do Sul de Minas Gerais**. EPAMIG, 2021. Disponível em: <https://livrariaepamig.com.br/wp-content/uploads/2023/02/ct-338.pdf>. Acesso em 02 jan. 2025.
- BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, n. 3, p. 345-370, 1987.
- BRESLOW, N. E. Discussion of professor Cox paper. **Jornal of the Royal statistical Society B**, v. 34, p. 216-217. 1972.
- BRESLOW, N. Covariance analysis of censored survival data. **Biometrics**, v. 30, n. 1, p. 89-99, 1974.
- CAIXETA, E. T.; RESENDE, M. D. V.; ALKIMIM, E. R.; SOUZA, T. V.; OLIVEIRA, A. C. B. D.; PEREIRA, A. A.; ALVES, R. S. **Aceleração do melhoramento do cafeeiro via seleção genômica: agilidade e eficácia no lançamento de novas cultivares**. EMBRAPA, 2022. Disponível em: <http://www.sbicafe.ufv.br/bitstream/handle/123456789/14342/Aceleracao-do-melhoramento-do-cafeeiro-Docmentos-17.pdf?sequence=1&isAllowed=y>. Acesso em 12 jan. 2025.

CARVALHO, C. H. S. de.; BARTELEGA, L.; SERA, G. H.; MATIELLO, J. B.; ALMEIDA, S. R. de; SANTINATO, F.; HOTZ, A. L. **Catálogo de cultivares de café arábica**. Embrapa. 2022.

CEMADEN. Mapa interativo. São José dos Campos: **Centro Nacional de Monitoramento e Alertas de Desastres Naturais**, 2025. Disponível em: <https://mapainterativo.cemaden.gov.br/#>. Acesso em: 02 jan. 2025.

CHALITA, L. V. A. S. **Modelos para dados agrupados e censurados**. 1997. 135 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1997.

CHALITA, L. V.; COLOSIMO, E. A.; DEMÉTRIO, C. G. B. Likelihood approximations and discrete models for tied survival data. **Communications in Statistics-Theory and Methods**, v. 31, n. 7, p. 1215-1229, 2002.

COLLETT, D. **Modelling survival data in medical research**. 2. ed. London. Chapman and Hall. 2023. 548 p.

COLOSIMO, E. A.; CHALITA, L. V.; DEMÉTRIO, C. G. B. Tests of proportional hazards and proportional odds models for grouped survival data. **Biometrics**, v. 56, n. 4, p. 1233-1240, 2000.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 1. ed. São Paulo, SP: Editora Blucher. 2006. 370 p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 2. ed. São Paulo, SP: Editora Blucher. 2024. 362 p.

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Safra de café encerra em 54,2 milhões de sacas em 2024 impactada por clima adverso**. Disponível em: <https://www.conab.gov.br/ultimas-noticias/5900-safra-de-cafe-encerra-em-54-2-milhoes-de-sacas-em-2024-impactada-por-clima-adverso>. Acesso em: 5 fev. 2025.

CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). **Café: formação da lavoura**. Brasília, 2016. Disponível em: [https://www.cnabrazil.org.br/assets/arquivos/188\\_Cafe\\_formacao\\_da\\_lavoura.pdf](https://www.cnabrazil.org.br/assets/arquivos/188_Cafe_formacao_da_lavoura.pdf). Acesso em: 5 fev. 2025.

COX, D. R. Partial likelihood. **Biometrika**, v. 62, n. 2, p. 269-276, 1975.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 30, n. 2, p. 248-265, 1968.

D'ARRIGO, G.; LEONARDIS, D.; ELHAFEEZ, S. A.; FUSARO, M.; TRIPEPI, G.; ROUMELIOTIS, S. Methods to analyse time-to-event data: the Kaplan-Meier survival curve. **Oxidative medicine and cellular longevity**, v. 2021, p. 1-7, 2021.

DUARTE, S. L.; PEREIRA, C. A.; TAVARES, M.; DOS REIS, E. A. Variáveis dos custos de produção versus preço de venda da cultura do café no segundo ano da lavoura. **REGE-Revista de Gestão**, v. 18, n. 4, p. 675-689, 2011.

EFRON, B. The efficiency of Cox's likelihood function for censored data. **Journal of the American statistical Association**, v. 72, n. 359, p. 557-565, 1977.

EFROYMSON, M. A. Multiple regression analysis. **Mathematical methods for digital computers**, p. 191-203, 1960.

EMILIANO, P. C.; VEIGA, E. P.; MENEZES, F. S. Critérios de informação de Akaike versus bayesiano: análise comparativa. **19º Simpósio Nacional de Probabilidade e Estatística**, 2010.

EMMERT-STREIB, F.; DEHMER, M. Introduction to survival analysis in practice. **Machine Learning and Knowledge Extraction**, v. 1, n. 3, p. 1013-1038, 2019.

FERREIRA, L. T.; CAVATON, F. T. **Safra dos cafés do Brasil foi estimada em 58,08 milhões de sacas para este ano de 2024**. Embrapa, 2024. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/86520314/artigo—safra-dos-cafes-do-brasil-foi-estimada-em-5808-milhoes-de-sacas-para-este-ano-de-2024>. Acesso em: 10 abr. 2024.

FILETTO, F.; ALENCAR, E. Introdução e expansão do café na região sul de Minas Gerais. **Organizações Rurais & Agroindustriais**, [S. l.], v. 3, n. 1, 2001. Disponível em: <https://www.revista.dae.ufla.br/index.php/ora/article/view/278>. Acesso em: 8 maio. 2024.

FONTES, J. R. M. **Heterose, capacidade combinatória e divergência genética estimada por análise de marcadores RAPD em cruzamentos entre cafeeiros Catuaí *Coffea arabica* L. e híbrido de Timor**. 2001. Tese de Doutorado. Universidade Federal de Viçosa. Viçosa, MG, 2001.

GIOLO, S. R. **Tópicos em análise de sobrevivência**. 2023. Disponível em: [https://docs.ufpr.br/~giolo/CE063/CodigosR/Comandos\\_TAS.html](https://docs.ufpr.br/~giolo/CE063/CodigosR/Comandos_TAS.html). Acesso em: 04 jan. 2025.

GUO, S. **Survival analysis**. 1. ed. New York, NY: Oxford University Press. 2010. 176 p.

JAYAWARDENE, S.; YANG, S. S. **A Nonlinear Ordered Rank Test to Detect Stochastic Ordering Between Two Distributions** 1. ed. New York, NY: Springer 1996. 334 p.

MARTINEZ, H. E. Prieto.; AUGUSTO, H.S.; CRUZ, C.D.; PEDROSA, A.W.; SAMPAIO, N.F. Crescimento vegetativo de cultivares de café (*Coffea arabica* L.) e sua correlação com a produção em espaçamentos adensados. **Acta Scientiarum. Agronomy**, v. 29, n. 4, p. 481-489, 2007.

- MARTINEZ, R. L. M. C.; NARANJO, J. D. Um pré-teste para escolher entre os testes de Logrank e Wilcoxon no problema de duas amostras. **Metron**, v. 68, p. 111-125, 2010.
- MIRANDA, J.M.; PERECIN, D.; PEREIRA, A.A. Produtividade e resistência à ferrugem do cafeeiro (*Hemileia vastatrix* Berk. Et. Br.) de progênies F5 de Catuaí Amarelo com o Híbrido de Timor. **Ciência e Agrotecnologia**, v.29, p.1195-1200, 2005.
- HERTZ-PICCIOTTO, I.; ROCKHILL, B. Validity and efficiency of approximation methods for tied survival times in Cox regression. **Biometrics**, p. 1151-1156, 1997.
- HOSMER, D. W.; LEMESHOW, S. **Applied Survival Analysis**, 1. ed. New York, NY: John Wiley and Sons, 1999. 392 p.
- HOSMER J. R., D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. 3. ed. Hoboken, John Wiley & Sons, 2013.
- INSTITUTO NACIONAL DE METEOROLOGIA, INMET. **Dados históricos**, 2025. Disponível em: <https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 20 jan. 2025.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the american statistical association**, v. 53, n. 282, p. 457-481, 1958.
- KLEINBAUM, D. G.; KLEIN, M. **Survival analysis a self-learning text**. 3. ed. New York, NY: Springer, 2012. 715 p.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. 1. ed. Nova Jersey, NJ: John Wiley e Sons, 2003. 664 p.
- LIMA, A. E. de.; SAMPAIO JUNIOR, H. G.; CASTRO, E. de M.; DE CARVALHO, S. P.; DA SILVA, F. F.; JÚNIOR, S. de L.; DE CARVALHO, A. M. INCIDENCE AND SEVERITY OF COFFEE LEAF RUST, CERCOSPORIOSIS AND COFFEE LEAF MINER IN COFFEE PROGENIES. **Coffee Science** - ISSN 1984-3909, [S. l.], v. 13, n. 3, p. 349-355, 2018. Disponível em: <https://coffeescience.ufla.br/index.php/Coffeescience/article/view/1459>. Acesso em: 25 apr. 2025.
- MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemother Rep**, v. 50, n. 3, p. 163-170, 1966.
- MARIZ, B. L.; CAIXETA, E. T.; RESENDE, M. D. V. D.; OLIVEIRA, A. C. B. D.; ALMEIDA, D. P. D.; ALVEZ, D. R. Exploring the Genetic Potential for Multi-Resistance to Rust and Other Coffee Phytopathogens in Breeding Programs. **Plants** v. 14, n. 3, p. 391, 2025.
- MCCOOK, S.; VANDERMEER, J. The big rust and the red queen: Long-term perspectives on coffee rust research. **Phytopathology**, v. 105, n. 9, p. 1164-1173, 2015.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction of theory of the statistics**. 3. ed. New York: John Wiley and Sons, 1974. 564 p

NAGAY, J. H. C. Café no Brasil: dois séculos de história. **UNICAMP, Formação Econômica**, Campinas, v. 3, n. 1, p. 17-23, 1999.

OLIVEIRA, A. C. B. D.; PEREIRA, A. A.; SILVA, F. L. D.; MENDONÇA, E. R. D.; FREITAS, M. A. S. D.; COELHO, I. F. **Produtividade de cultivares de café arábica em duas regiões cafeeiras de Minas Gerais**. 2012. Disponível em: <http://www.sbicafe.ufv.br/handle/123456789/5149>. Acesso: 26 de dez 2024.

OYAMA F.; WILSON. K.; SANTANA, S. H. S. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 7, n. 1, 2020.

Pereira, Eliseu J.G.; Picanço, Marcelo C.; Guedes, Raul N.C.; Fernandes, Flávio L.; Crespo, André L.B.; Rosado, Jander F. Controle natural do bicho mineiro do cafeeiro no início do período seco. In: Simpósio de Pesquisa dos Cafés do Brasil e Workshop Internacional de Café Saúde, (3. : 2003 : Porto Seguro). Anais. Brasília, DF : Embrapa Café, 2003. (447p.), p. 330-331.

PAULO, E.M.; JUNIOR, E.F.; FAZUOLI, L.C. Comportamento de cultivares de cafeeiro em diferentes densidades de plantio. **Bragantia**, v.64, p.397-409, 2005.

PAVAN, M.A.; CHAVES, J.C.D.; ANDROCIOLI-FILHO, A. Produção de café (*Coffea arabica* L.) var. Caturra em função da densidade de plantio, adubação e tratamento fitossanitário. **Turrialba**, v.44, p.227-231, 1994.

PETO, R. Contribution to the discussion of Cox. **Jornal of the Royal Statistical society**, v. 34, p. 205-207, 1972.

POMA-ANGAMARCA, R. A.; ROJAS, J. R.; SÁNCHEZ-RODRÍGUEZ, A.; RUÍZ-GONZÁLEZ, M. X. Diversity of Leaf Fungal Endophytes from Two *Coffea arabica* Varieties and Antagonism towards Coffee Leaf Rust. **Plants**, v. 13, n. 6, p. 814, 2024.

PRENTICE, R. L.; GLOECKLER, L. A. Regression analysis of grouped survival data with application to breast cancer data. **Biometrics**, v. 34, p. 57-67, 1978.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. 2025. Disponível em: <<https://www.R-project.org/>>.

SCHOENFELD, D. Partial residuals for the proportional hazard regression model. **Biometrika**, v. 69, p. 239-241, 1982.

SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, p. 461-464, 1978.

SOBRAL, T. E. L.; BARRETO, G. Análise dos critérios de informação para a seleção de ordem em modelos auto-regressivos. In: **X Conferência Brasileira de Dinâmica, Controle e Aplicações**. 2011.

STRAPASSON, E. **Comparação de modelos com censura intervalar em análise de sobrevivência**. 2007. Tese de Doutorado. Universidade de São Paulo. São Paulo, SP, 2007.

TURKSON, A. J.; AYIAH-MENSAH, F.; NIMOH, V. Handling censoring and censored data in survival analysis: A standalone systematic literature review. **International journal of mathematics and mathematical sciences**, v. 2021, p. 1-16, 2021.

THERNEAU, T. M.; GRAMBSCH, P. M.; FLEMING, T. R. Martingale-based residuals for survival models. **Biometrika**, v. 77, n. 1, p. 147-160, 1990.

VRIEZE, S. I. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). **Psychological methods**, v. 17, n. 2, p. 228, 2012.

## APÊNDICE

	<b>Página</b>
APÊNDICE A: Resumo dos dados. ....	<a href="#">101</a>
APÊNDICE B: Rotina análise dos dados. ....	<a href="#">102</a>
APÊNDICE C: Rotina para seleção de variáveis. ....	<a href="#">121</a>

## APÊNDICE A: Resumo dos dados.

# Resumo dos dados utilizados até o processo das aproximações de verossimilhança parcial.

Dados utilizados na comparação com a testemunha Paraíso

# A tibble: 6 × 27

	planta	STAT	ANO	CENS	PREC	TEMP	TEMPO	FIL	BLO	Trat	P1	VIG	PRO	TF	CMT
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	7	0	2022	0	1621.	19	6	1	1	22	1	7	0	3	5
2	8	0	2016	3	1343.	20.7	0	1	1	22	2	0	0	0	0
3	9	1	2020	1	1062	21.3	4	1	1	22	3	7	4	3	3
4	10	1	2020	1	1062	21.3	4	1	1	22	4	7.5	5	3	3
5	11	1	2020	1	1062	21.3	4	1	1	22	5	7.5	4.5	4	2
6	12	1	2020	1	1062	21.3	4	1	1	22	6	7	3.5	4	3

# 12 more variables: UMT <dbl>, CER <dbl>, FER <dbl>, BM <dbl>, APL <dbl>, DCO <dbl>, DCA1 <dbl>, DCA <dbl>, NRP <dbl>, CRP1 <dbl>, CRP <dbl>, NNR <dbl>

Resumo dos dados utilizados para análise dos modelos discretos em comparação com a testemunha Paraíso

# A tibble: 6 × 32

	Obs	int1	int2	int3	int4	int5	int	y	planta	stat	ANO	PREC	TEMP	TIME	FIL
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	1	0	0	0	0	1	0	7	0	2022	1621.	19	6	1
2	2	0	1	0	0	0	2	0	7	0	2022	1621.	19	6	1
3	3	0	0	1	0	0	3	0	7	0	2022	1621.	19	6	1
4	4	0	0	0	1	0	4	0	7	0	2022	1621.	19	6	1
5	5	0	0	0	0	1	5	0	7	0	2022	1621.	19	6	1
6	6	1	0	0	0	0	1	0	8	0	2016	1343.	20.7	0	1

# 17 more variables: BLO <dbl>, Trat <dbl>, P1 <dbl>, VIG <dbl>, PRO <dbl>, TF <dbl>, CMT <dbl>, UMT <dbl>, CER <dbl>, FER <dbl>, BM <dbl>, APL <dbl>, DCO <dbl>, DCA <dbl>, NRP <dbl>, CRP <dbl>, NNR <dbl>

[1] "Obs" "int1" "int2" "int3" "int4" "int5" "int" "y" "planta" "stat"

[11] "ANO" "PREC" "TEMP" "TIME" "FIL" "BLO" "Trat" "P1" "VIG" "PRO"

[21] "TF" "CMT" "UMT" "CER" "FER" "BM" "APL" "DCO" "DCA" "NRP"

[31] "CRP" "NNR"

Dados utilizados na comparação com a testemunha Catuaí Vermelho

# A tibble: 6 × 24

	planta	stat	ANO	PREC	TEMP	TIME	FIL	BLO	Trat	P1	VIG	TF	CMT	UMT	PRO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	7	0	2022	1621.	19	6	1	1	22	1	7	3	5	1	0.1
2	8	0	2016	1343.	20.7	0	1	1	22	2	0	0	0	0	0
3	9	1	2020	1062	21.3	4	1	1	22	3	7	3	3	3	4
4	10	1	2020	1062	21.3	4	1	1	22	4	7.5	3	3	2	5
5	11	1	2020	1062	21.3	4	1	1	22	5	7.5	4	2	2	4.5
6	12	1	2020	1062	21.3	4	1	1	22	6	7	4	3	3	3.5

# 9 more variables: CER <dbl>, FER <dbl>, BM <dbl>, APL <dbl>, DCO <dbl>, DCA <dbl>, NRP <dbl>, CRP <dbl>, NNR <dbl>

Dados utilizados na abordagem discreta, na comparação realizada com a testemunha Catuaí vermelho:

# A tibble: 6 × 29

	Obs	int1	int2	int3	int4	int5	int	y	planta	STAT	ANO	PREC	Temp	TIME	FIL
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	1	0	0	0	0	1	0	7	0	2022	1621.	19	6	1
2	2	0	1	0	0	0	2	0	7	0	2022	1621.	19	6	1
3	3	0	0	1	0	0	3	0	7	0	2022	1621.	19	6	1
4	4	0	0	0	1	0	4	0	7	0	2022	1621.	19	6	1
5	5	0	0	0	0	1	5	0	7	0	2022	1621.	19	6	1
6	6	1	0	0	0	0	1	0	8	0	2016	1343.	20.7	0	1

# 14 more variables: BLO <dbl>, Trat <dbl>, P1 <dbl>, VIG <dbl>, PRO <dbl>, CER <dbl>, FER <dbl>, BM <dbl>, APL <dbl>, DCO <dbl>, DCA <dbl>, NRP <dbl>, CRP <dbl>, NNR <dbl>

## APÊNDICE B: Rotina para análise de dados.

```
#####
##### Comparação com a testemunha Paraíso #####
#####

rm(list=ls())
ls()

#### Carregando os dados
library(readxl)
Testemunha1 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/Testemunha1.xlsx")
head(Testemunha1); attach(Testemunha1)

# Paleta de cores para montagem dos gráficos
# Criando o data frame com nomes e códigos das cores
cores_df <- data.frame(
  Nome = c("Vermelho", "Azul", "Amarelo", "Laranja", "Verde", "Roxo",
           "Ciano", "Magenta", "Lima", "Vermelho-alaranjado"),
  Codigo_Hex = c("#FF0000", "#0000FF", "#FFFF00", "#FFA500", "#008000", "#800080",
                 "#00FFFF", "#FF00FF", "#BFFF00", "#FF4500")
)

# Exibindo o data frame
print(cores_df)

library(survival)
library(survminer)
# Curvas de sobrevivência utilizando o estimador KM
# Testemunha 1

KM<-survfit(Surv(time,stat)~1,data = Testemunha1);KM
summary(KM)
surv_summary(KM)

# Gráfico da curva de produção via KM

par(mar = c(4, 4, 1, 2)) # Margens personalizadas
plot(KM, ylim = c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
     fun="event", xaxt="n", yaxt="n", lwd=3, main="")
axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

# Formatando os valores do eixo Y para exibir vírgula como separador decimal
eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

# Ajuste de Kaplan-Meier para diferentes blocos
fit1 <- survfit(Surv(time, stat) ~ Bl, data = Testemunha1)
summary(fit1)

#rho = 0 implica no teste log-rank, rho = 1 teste de Wilcoxon

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ Bl, data = Testemunha1, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ Bl, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

#library(survminer)
pairwise_survdiff(Surv(time, stat) ~ Bl, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ Bl, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

survdiff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("19","20")), rho=1)
survdiff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("19","21")), rho=1)
```

```

survdifff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("19","22")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("20","21")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("20","22")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha1, Bl %in% c("21","22")), rho=1)

# Gráfico de produção acumulada nos Blocos

{
par(mar = c(4, 4, 1, 2)) # Margens personalizadas
lty<-c(1,2,1,2)
col<- c(1:4)
plot(fit1, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
      fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

# Formatando os valores do eixo Y para exibir vírgula como separador decimal
eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

legend(0,0.9, lty=lty,col=col, c("Bloco 1", "Bloco 2", "Bloco 3",
                                "Bloco 4"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos das fileiras - Testemunha1
# Ajuste de Kaplan-Meier para diferentes fileiras
fit2 <- survfit(Surv(time, stat) ~ Fil, data = Testemunha1)
summary(fit2)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

#library(survminer)
#pairwise_survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
#pairwise_survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

# Gráfico de produção acumulada das fileiras

{
par(mar = c(4, 4, 0, 2)) # Margens personalizadas
lty<-c(1,1,1,1,2)
col<- c(col <- c("#1B9E77", "#D95F02", "#7570B3", "#E7298A", "#66A61E")
)
plot(fit2, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
      fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

# Formatando os valores do eixo Y para exibir vírgula como separador decimal
eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

legend(0,0.9, lty=lty,col=col, c("Fileira 1", "Fileira 2", "Fileira 3",
                                "Fileira 4", "Fileira 5"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de Ferrugem - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de Ferrugem
fit3 <- survfit(Surv(time, stat) ~ FER, data = Testemunha1)
summary(fit3)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ FER, data = Testemunha1, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ FER, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdifff(Surv(time, stat) ~ FER, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdifff(Surv(time, stat) ~ FER, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

```

```

survdif(Surv(time,stat)~FER, data=subset(Testemunha1, FER %in% c("1","2")), rho=1)

# Gráfico relativo a Ferrugem
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(1,1)
  col<- c("#1F78B4", "#000000")
  plot(fit3, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0,0.9, lty=lty,col=col, c("Nível 1", "Nível 2"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de BM - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de BM

fit4 <- survfit(Surv(time, stat) ~ BM, data = Testemunha1)
summary(fit4)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ BM, data = Testemunha1, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ BM, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdif(Surv(time, stat) ~ BM, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdif(Surv(time, stat) ~ BM, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

# Gráfico relativo a Bicho Mineiro
{
  par(mar = c(4, 4, 0, 2)) # Margens personalizadas
  lty<-c(2,1,1)
  col<- c("#0000FF", "#000000","#FFA500" )
  plot(fit4, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Folhas com poucas e pequenas lesões", "Imune","Folhas com poucas lesões"),
        lwd = 3, bty = "n", cex = 0.8) # Reduz o tamanho da legenda
}

survdif(Surv(time,stat)~BM, data=subset(Testemunha1, BM %in% c("1","2")), rho=0)
survdif(Surv(time,stat)~BM, data=subset(Testemunha1, BM %in% c("1","3")), rho=0)
survdif(Surv(time,stat)~BM, data=subset(Testemunha1, BM %in% c("2","3")), rho=0)

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de Cercospora - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de CER - Cercospora

fit5 <- survfit(Surv(time, stat) ~ CER, data = Testemunha1)
summary(fit5)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ CER, data = Testemunha1, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ CER, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdif(Surv(time, stat) ~ CER, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdif(Surv(time, stat) ~ CER, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

```

```

# Gráfico da cercospora
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,2,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000" )
  plot(fit5, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Moderada para muita","Imune", "Pouca", "Pouca para média"),
        lwd = 3, bty = "n", cex = 1.0) # Reduz o tamanho da legenda
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos da Uniformidade de maturação - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de UMT

fit6 <- survfit(Surv(time, stat) ~ UMT, data = Testemunha1)
summary(fit6)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ UMT, data = Testemunha1);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ UMT, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ UMT, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ UMT, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

# Gráfico da Uniforme dos frutos
{
  par(mar = c(4, 4, 0, 2)) # Margens personalizadas
  lty<-c(2,1,2,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000" )
  plot(fit6, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Desuniforme", "Uniforme", "Medianamente uniforme", "Medianamente desuniforme"),
        lwd = 3, bty = "n", cex = 1.0) # Reduz o tamanho da legenda
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos do Tamanho do fruto - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de TF-Tamanho do fruto

fit7 <- survfit(Surv(time, stat) ~ TF, data = Testemunha1)
summary(fit7)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ TF, data = Testemunha1);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ TF, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ TF, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ TF, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

```

```

# Gráfico referente ao tamanho do fruto
{
  par(mar = c(4, 4, 0, 2)) # Margens personalizadas
  lty<-c(2,1,2,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000" )
  plot(fit7, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("grande","miúdo", "pequeno","médio"),
        lwd = 3, bty = "n", cex = 1.0) # Reduz o tamanho da legenda
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos ciclo de maturação - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de CMT - Ciclo de maturação

fit8 <- survfit(Surv(time, stat) ~ CMT, data = Testemunha1)
summary(fit8)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ CMT, data = Testemunha1, rho=0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ CMT, data = Testemunha1, rho = 1);Wilcox_test
p_value <- 1-pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ CMT, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ CMT, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

# Gráfico do ciclo de maturação
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,2,1,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000", "#000")
  plot(fit8, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("tardia", "precoce", "media para precoce", "media","media para tardia"),
        lwd = 2, bty = "n", cex = 0.8) # Reduz o tamanho da legenda
}

survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","2")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","3")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","4")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","5")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","3")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","4")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","5")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("3","4")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("3","5")), rho=0)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("4","5")), rho=0)

survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","2")), rho=1)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","3")), rho=1)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","4")), rho=1)
survdiff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("1","5")), rho=1)

```

```

survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","3")), rho=1)
survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","4")), rho=1)
survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("2","5")), rho=1)
survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("3","4")), rho=1)
survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("3","5")), rho=1)
survdifff(Surv(time,stat)~CMT, data=subset(Testemunha1, CMT %in% c("4","5")), rho=1)

### Log - rank para as covariáveis categóricas
logrank_test <-survdifff(Surv(time, stat) ~ Bl, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ FER, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ BM, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ CER, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ UMT, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ CMT, data = Testemunha1, rho=0);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ TF, data = Testemunha1, rho=0);logrank_test

### Wilcox para as variáveis categóricas
logrank_test <-survdifff(Surv(time, stat) ~ Bl, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ FER, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ BM, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ CER, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ UMT, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ CMT, data = Testemunha1, rho=1);logrank_test
logrank_test <-survdifff(Surv(time, stat) ~ TF, data = Testemunha1, rho=1);logrank_test

pairwise_survdifff(Surv(time, stat) ~ CMT, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdifff(Surv(time, stat) ~ CMT, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

rm(list=ls())
ls()

library(readxl)
Testemunha1 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/Testemunha1.xlsx")
head(Testemunha1); attach(Testemunha1)

library(survival)

# Modelo de Cox
#Aproximação Breslow
Modelofinal_BRESLOW1<-coxph(Surv(time,stat)~ APL+DCA+Vig, data = Testemunha1, ties = "breslow")
summary(Modelofinal_BRESLOW1); anova(Modelofinal_BRESLOW1)

#Verificando a pressuposição dos residuo de schenfeld
resid(Modelofinal_BRESLOW1, type ="scaledsch")
cox.zph(Modelofinal_BRESLOW1, transform = "rank", terms = TRUE, global = TRUE)
par(mfrow=c(1,3))
plot(cox.zph(Modelofinal_BRESLOW1))

# Ajusta margens
par(mar = c(4, 4, 1, 2))

# Calcula resíduos de martingale
resm <- resid(Modelofinal_BRESLOW1, type = "martingale")
res <- Testemunha1$stat - resm

# Ajusta modelo de sobrevivência com os resíduos
ekm <- survfit(Surv(res, Testemunha1$stat) ~ 1)

# Define separador decimal como vírgula
options(OutDec = ",")

# Plota gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico 1: Curva Kaplan-Meier e curva exponencial teórica
plot(ekm, mark.time = FALSE, conf.int = FALSE,

```

```

        xlab = "Resíduos", ylab = "S(e) estimada")
res_ord <- sort(res)
exp1 <- exp(-res_ord)
lines(res_ord, exp1, lty = 3)
legend("topright", legend = c("Kaplan-Meier", "Exponencial padrão"),
      lty = c(1, 3), lwd = 1, bty = "n", cex = 0.6)

# Gráfico 2: S(e) observado vs S(e) teórico
st <- ekm$surv
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1,
      xlab = "S(e): Kaplan-Meier", ylab = "S(e): Exponencial padrão",
      pch = 16)
abline(a = 0, b = 1, lty = 2, col = "black") # Linha de 45 graus

#Aproximação Efron
Modelofinal_EFRON1<-coxph(Surv(time,stat)~DCA+Vig+NRP+temper, data = Testemunha1, ties = "efron")
summary(Modelofinal_EFRON1); anova(Modelofinal_EFRON1)

#Verificando a pressuposição dos residuo de schenfeld
resid(Modelofinal_EFRON1, type = "scaledsch")
cox.zph(Modelofinal_EFRON1, transform = "identity", terms = TRUE, global = TRUE)
plot(cox.zph(Modelofinal_EFRON1))

# Ajusta margens
par(mar = c(4, 4, 1, 2))

# Calcula resíduos de martingale
resm <- resid(Modelofinal_EFRON1, type = "martingale")
res <- Testemunha1$stat - resm

# Ajusta modelo de sobrevivência com os resíduos
ekm <- survfit(Surv(res, Testemunha1$stat) ~ 1)

# Define separador decimal como vírgula
options(OutDec = ",")

# Plota gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico 1: Curva Kaplan-Meier e curva exponencial teórica
plot(ekm, mark.time = FALSE, conf.int = FALSE,
      xlab = "Resíduos", ylab = "S(e) estimada")
res_ord <- sort(res)
exp1 <- exp(-res_ord)
lines(res_ord, exp1, lty = 3)
legend("topright", legend = c("Kaplan-Meier", "Exponencial padrão"),
      lty = c(1, 3), lwd = 1, bty = "n", cex = 0.6)

# Gráfico 2: S(e) observado vs S(e) teórico
st <- ekm$surv
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1,
      xlab = "S(e): Kaplan-Meier", ylab = "S(e): Exponencial padrão",
      pch = 16)
abline(a = 0, b = 1, lty = 2, col = "black") # Linha de 45 graus

## Modelo de aproximações exact

Modelofinal_EXACT1<-coxph(Surv(time,stat)~NNR+Vig+temper+APL, data = Testemunha1, ties = "exact")
summary(Modelofinal_EXACT1); anova(Modelofinal_EXACT1)
mf<-2*as.numeric(logLik(Modelofinal_EXACT1));mf

#Verificando a pressuposição dos residuo de schenfeld

```

```

resid(Modelofinal_EXACT1, type = "scaledsch")
cox.zph(Modelofinal_EXACT1, transform = "identity", terms = TRUE, global = TRUE)
plot(cox.zph(Modelofinal_EXACT1))

# Ajusta margens
par(mar = c(4, 4, 1, 2))

# Calcula resíduos de martingale
resm <- resid(Modelofinal_EXACT1, type = "martingale")
res <- Testemunha1$stat - resm

# Ajusta modelo de sobrevivência com os resíduos
ekm <- survfit(Surv(res, Testemunha1$stat) ~ 1)

# Define separador decimal como vírgula
options(OutDec = ",")

# Plota gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico 1: Curva Kaplan-Meier e curva exponencial teórica
plot(ekm, mark.time = FALSE, conf.int = FALSE,
     xlab = "Resíduos", ylab = "S(e) estimada")
res_ord <- sort(res)
exp1 <- exp(-res_ord)
lines(res_ord, exp1, lty = 3)
legend("topright", legend = c("Kaplan-Meier", "Exponencial padrão"),
      lty = c(1, 3), lwd = 1, bty = "n", cex = 0.6)

# Gráfico 2: S(e) observado vs S(e) teórico
st <- ekm$surv
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1,
     xlab = "S(e): Kaplan-Meier", ylab = "S(e): Exponencial padrão",
     pch = 16)
abline(a = 0, b = 1, lty = 2, col = "black") # Linha de 45 graus

## Dados agrupados
rm(list=ls())
ls()

library(readxl)
intervalos_t1 <- read_excel("Mestrado/DISSERTAÇÃO/Base de dados/intervalos_T1.xlsx")
head(intervalos_t1); names(intervalos_t1)

# Modelo de riscos proporcionais de Cox
modelo_um <- glm(formula = y ~ Temp + int1 + int2 + APL + int4 + Vig +
                 NNR, family = binomial(link = "cloglog"), data = intervalos_t1)

summary(modelo_um); anova(modelo_um)

# Curva ROC
{
require(ROCR)

# Ajustar separador decimal global
options(OutDec = ",")

# Predição e performance
pred <- prediction(modelo_um$fitted.values, y)
perf <- performance(pred, "tpr", "fpr")
auc <- performance(pred, "auc")@y.values[[1]]

#par(mar = c(inferior, esquerda, superior, direita))
par(mar = c(5, 5, 4, 2))

```

```

# Plot vazio (sem eixos) para customização
plot(perf,
      col = "blue",
      lwd = 2,
      main = "Curva ROC",
      xlab = "Taxa de falsos positivos (FPR)",
      ylab = "Taxa de verdadeiros positivos (TPR)",
      cex.lab = 1.2,
      cex.main = 1.4,
      xaxt = "n", # remove eixo X padrão
      yaxt = "n") # remove eixo Y padrão

# Customizar eixos com separador vírgula
axis(1, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))
axis(2, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))

# Linha diagonal de referência
abline(a = 0, b = 1, lty = 2, col = "black")

# Grid
grid()

# Exibir AUC no gráfico com vírgula
legend("bottomright",
      legend = paste("AUC =", format(round(auc, 3), decimal.mark = ",")),
      bty = "n")

# Exibir AUC no console
auc

}

# Modelo Logístico Testemunha 1
modelo_dois <- glm(formula = y ~ Temp + int1 + int2 + APL + int4 + Vig +
                  NNR, family = binomial(link = "logit"), data = intervalos_t1)

AIC(modelo_dois); BIC(modelo_dois)

summary(modelo_dois); anova(modelo_dois)

# Curva ROC
{
require(ROCR)

# Ajustar separador decimal global
options(OutDec = ",")

# Predição e performance
pred <- prediction(modelo_dois$fitted.values, intervalos_t1$y)
perf <- performance(pred, "tpr", "fpr")
auc <- performance(pred, "auc")@y.values[[1]]

#par(mar = c(inferior, esquerda, superior, direita))
par(mar = c(4, 4, 1, 1))

# Plot vazio (sem eixos) para customização
plot(perf,
      col = "blue",
      lwd = 2,
      main = "",
      xlab = "Taxa de falsos positivos",
      ylab = "Taxa de verdadeiros positivos",
      cex.lab = 1.2,
      cex.main = 1.4,
      xaxt = "n", # remove eixo X padrão
      yaxt = "n") # remove eixo Y padrão

```

```

# Customizar eixos com separador vírgula
axis(1, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))
axis(2, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))

# Linha diagonal de referência
abline(a = 0, b = 1, lty = 2, col = "black")

# Grid
grid()

# Exibir AUC no gráfico com vírgula
legend("bottomright",
      legend = paste("AUC =", format(round(auc, 4), decimal.mark = ",")),
      bty = "n")

# Exibir AUC no console
auc
}

#####
#####          Comparação com a testemunha Catuaí Vermelho          #####
#####

rm(list=ls())
ls()

#### Carregando os dados

library(readxl)
Testemunha2 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/Testemunha_2.xlsx")
head(Testemunha2); attach(Testemunha2)

library(survival)
library(survminer)

# Curvas de sobrevivência utilizando o estimador KM

KM<-survfit(Surv(time,stat)~1,data = Testemunha2);KM
summary(KM)
surv_summary(KM)

par(mar = c(4, 4, 1, 2)) # Margens personalizadas
plot(KM, ylim = c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
     fun="event", xaxt="n", yaxt="n", lwd=3, main="")
axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

# Formatando os valores do eixo Y para exibir vírgula como separador decimal
eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

# Ajuste de Kaplan-Meier para diferentes blocos
fit1 <- survfit(Surv(time, stat) ~ B1, data = Testemunha2)
summary(fit1)

#rho = 0 implica no teste log-rank, rho = 1 teste de Wilcoxon

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ B1, data = Testemunha2, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ B1, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

#library(survminer)
pairwise_survdiff(Surv(time, stat) ~ B1, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ B1, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

```

```

survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","20")), rho=0)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","21")), rho=0)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","22")), rho=0)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("20","21")), rho=0)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("20","22")), rho=0)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("21","22")), rho=0)

survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","20")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","21")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("19","22")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("20","21")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("20","22")), rho=1)
survdifff(Surv(time,stat)~Bl, data=subset(Testemunha2, Bl %in% c("21","22")), rho=1)

# Gráfico de produção acumulada nos Blocos

{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(1,2,1,2)
  col<- c(1:4)
  plot(fit1, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0,0.9, lty=lty,col=col, c("Bloco 1", "Bloco 2", "Bloco 3",
                                   "Bloco 4"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos das fileiras - Testemunha1

# Ajuste de Kaplan-Meier para diferentes fileiras
fit2 <- survfit(Surv(time, stat) ~ Fil, data = Testemunha2)
summary(fit2)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ Fil, data = Testemunha2, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ Fil, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

#library(survminer)
#pairwise_survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 0, p.adjust.method = "bonferroni")
#pairwise_survdifff(Surv(time, stat) ~ Fil, data = Testemunha1, rho = 1, p.adjust.method = "bonferroni")

# Gráfico de produção acumulada das fileiras
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(1,1,1,1,2)
  col<- c(col <- c("#1B9E77", "#D95F02", "#7570B3", "#E7298A", "#66A61E")
)
  plot(fit2, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0,0.9, lty=lty,col=col, c("Fileira 1", "Fileira 2", "Fileira 3",
                                   "Fileira 4", "Fileira 5"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de Ferrugem - Testemunha1

# Ajuste de Kaplan-Meier para diferentes níveis de Ferrugem

```

```

fit3 <- survfit(Surv(time, stat) ~ FER, data = Testemunha2)
summary(fit3)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ FER, data = Testemunha2, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ FER, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ FER, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ FER, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

survdiff(Surv(time,stat)~FER, data=subset(Testemunha2, FER %in% c("1","2")), rho=0)

# Gráfico relativo a Ferrugem
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(1,1)
  col<- c("#1F78B4", "#000000")
  plot(fit3, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0,0.9, lty=lty,col=col, c( "Nível 2", "Nível 1"), lwd=3, bty="n")
}

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de BM - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de BM

fit4 <- survfit(Surv(time, stat) ~ BM, data = Testemunha2)
summary(fit4)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ BM, data = Testemunha2, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ BM, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ BM, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ BM, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

# Gráfico relativo a Bicho Mineiro
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,1)
  col<- c("#0000FF", "#000000","#FFA500" )
  plot(fit4, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Folhas com poucas e pequenas lesões", "Imune","Folhas com poucas lesões"),
        lwd = 2, bty = "n", cex = 0.7) # Reduz o tamanho da legenda
}

survdiff(Surv(time,stat)~BM, data=subset(Testemunha2, BM %in% c("1","2")), rho=1)
survdiff(Surv(time,stat)~BM, data=subset(Testemunha2, BM %in% c("1","3")), rho=1)
survdiff(Surv(time,stat)~BM, data=subset(Testemunha2, BM %in% c("2","3")), rho=1)

# Ajuste dos gráficos de sobrevivência para analisar os efeitos de Cercospora - Testemunha1

```

```

# Ajuste de Kaplan-Meier para diferentes níveis de CER - Cercospora

fit5 <- survfit(Surv(time, stat) ~ CER, data = Testemunha2)
summary(fit5)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ CER, data = Testemunha2, rho = 0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ CER, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ CER, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ CER, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

# Gráfico da cercospora
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,2,1)
  col<- c("#008000", "#FFA500", "#FF0000" )
  plot(fit5, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Pouca para média", "Sem cercóspora", "Pouca "),
        lwd = 2, bty = "n", cex = 0.7) # Reduz o tamanho da legenda
}

survdiff(Surv(time,stat)~CER, data=subset(Testemunha2_filtrado, CER %in% c("1","2")), rho=0)
survdiff(Surv(time,stat)~CER, data=subset(Testemunha2_filtrado, CER %in% c("1","3")), rho=0)
survdiff(Surv(time,stat)~CER, data=subset(Testemunha2_filtrado, CER %in% c("2","3")), rho=0)

# Ajuste dos gráficos de sobrevivência para analisar os efeitos da Uniformidade de maturação - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de UMT

fit6 <- survfit(Surv(time, stat) ~ UMT, data = Testemunha2)
summary(fit6)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ UMT, data = Testemunha2);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ UMT, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdiff(Surv(time, stat) ~ UMT, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdiff(Surv(time, stat) ~ UMT, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

# Gráfico da Uniforme dos frutos
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,2,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000" )
  plot(fit6, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("Desuniforme", "Uniforme", "Medianamente uniforme", "Medianamente desuniforme"),
        lwd = 2, bty = "n", cex = 0.7) # Reduz o tamanho da legenda
}

```

```

survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","2")), rho=0)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","3")), rho=0)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","4")), rho=0)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("2","3")), rho=0)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("2","4")), rho=0)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("3","4")), rho=0)

survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","2")), rho=1)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","3")), rho=1)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("1","4")), rho=1)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("2","3")), rho=1)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("2","4")), rho=1)
survdif(Surv(time,stat)~UMT, data=subset(Testemunha2, UMT %in% c("3","4")), rho=1)

# Ajuste dos gráficos de sobrevivência para analisar os efeitos do Tamanho do fruto - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de TF-Tamanho do fruto

fit7 <- survfit(Surv(time, stat) ~ TF, data = Testemunha2)
summary(fit7)

# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ TF, data = Testemunha2);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ TF, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1 - pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

pairwise_survdif(Surv(time, stat) ~ TF, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdif(Surv(time, stat) ~ TF, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

# Gráfico referente ao tamanho do fruto

{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,1)
  col<- c("#008000", "#FFA500", "#FF0000" )
  plot(fit7, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=3, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("grande", "pequeno", "médio"),
        lwd = 2, bty = "n", cex = 1.0) # Reduz o tamanho da legenda
}

survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("2","3")), rho=0)
survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("2","4")), rho=0)
survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("3","4")), rho=0)

survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("2","3")), rho=1)
survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("2","4")), rho=1)
survdif(Surv(time,stat)~TF, data=subset(Testemunha2, TF %in% c("3","4")), rho=1)

# Ajuste dos gráficos de sobrevivência para analisar os efeitos ciclo de maturação - Testemunha1
# Ajuste de Kaplan-Meier para diferentes níveis de CMT - Ciclo de maturação

fit8 <- survfit(Surv(time, stat) ~ CMT, data = Testemunha2)
summary(fit8)
CMT
# Teste de log-rank para comparação entre grupos
logrank_test <- survdiff(Surv(time, stat) ~ CMT, data = Testemunha2, rho=0);logrank_test
Wilcox_test <- survdiff(Surv(time, stat) ~ CMT, data = Testemunha2, rho = 1);Wilcox_test
p_value <- 1-pchisq(logrank_test$chisq, df = length(logrank_test$n) - 1)

```

```

pairwise_survdif(Surv(time, stat) ~ CMT, data = Testemunha2, rho = 0, p.adjust.method = "bonferroni")
pairwise_survdif(Surv(time, stat) ~ CMT, data = Testemunha2, rho = 1, p.adjust.method = "bonferroni")

# Gráfico do cilco de maturação
{
  par(mar = c(4, 4, 1, 2)) # Margens personalizadas
  lty<-c(2,1,2,1,1)
  col<- c("#0000FF", "#008000", "#FFA500", "#FF0000", "#000")
  plot(fit8, lty=lty,col=col, ylim=c(0,1), xlab= "Tempo (anos)", ylab = "Probabilidade de produção",
       fun="event",xaxt="n",yaxt="n", lwd=2, main="") #xaxt para não soltar os números do eixo x
  axis(1, xaxp=c(0,6,6)) # Intervalo e subdivisões do eixo X

  # Formatando os valores do eixo Y para exibir vírgula como separador decimal
  eixo_y <- seq(0, 1, length.out = 11) # Gera os valores do eixo Y
  axis(2, at = eixo_y, labels = format(eixo_y, decimal.mark=",")) # Aplica formatação

  legend(0, 0.9, lty = lty, col = col,
        legend = c("precoce", "tardia", "média", "média para precoce", "média para tardia"),
        lwd = 2, bty = "n", cex = 0.7) # Reduz o tamanho da legenda
}

survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","2")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","3")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","4")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","5")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","3")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","4")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","5")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("3","4")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("3","5")), rho=0)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("4","5")), rho=0)

survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","2")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","3")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","4")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("1","5")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","3")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","4")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("2","5")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("3","4")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("3","5")), rho=1)
survdif(Surv(time,stat)~CMT, data=subset(Testemunha2, CMT %in% c("4","5")), rho=1)

### Log - rank para as covariáveis categóricas
logrank_test <-survdif(Surv(time, stat) ~ Bl, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ Fil, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ FER, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ BM, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ CER, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ UMT, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ CMT, data = Testemunha2, rho=0);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ TF, data = Testemunha2, rho=0);logrank_test

### Wilcox para as variáveis categóricas
logrank_test <-survdif(Surv(time, stat) ~ Bl, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ Fil, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ FER, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ BM, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ CER, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ UMT, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ CMT, data = Testemunha2, rho=1);logrank_test
logrank_test <-survdif(Surv(time, stat) ~ TF, data = Testemunha2, rho=1);logrank_test

rm(list=ls())
ls()

```

```

library(readxl)
Testemunha2 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/Testemunha_2.xlsx")
head(Testemunha2); attach(Testemunha2)

#### Modelo de Cox ####
# Aproximação de Breslow
Modelofinal_BRESLOW2<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = Testemunha2, ties = "efron")
summary(Modelofinal_BRESLOW2); anova(Modelofinal_BRESLOW2)

#Verificando a pressuposição dos resíduo de schenfeld
resid(Modelofinal_BRESLOW2, type = "scaledsch")
cox.zph(Modelofinal_BRESLOW2, transform = "identity")
par(mfrow=c(2,2))
plot(cox.zph(Modelofinal_BRESLOW2))

# Ajusta margens
par(mar = c(4, 4, 1, 2))

# Calcula resíduos de martingale
resm <- resid(Modelofinal_BRESLOW2, type = "martingale")
res <- Testemunha2$stat - resm

# Ajusta modelo de sobrevivência com os resíduos
ekm <- survfit(Surv(res, Testemunha2$stat) ~ 1)

# Define separador decimal como vírgula
options(OutDec = ",")

# Plota gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico 1: Curva Kaplan-Meier e curva exponencial teórica
plot(ekm, mark.time = FALSE, conf.int = FALSE,
     xlab = "Resíduos", ylab = "S(e) estimada")
res_ord <- sort(res)
exp1 <- exp(-res_ord)
lines(res_ord, exp1, lty = 3)
legend("topright", legend = c("Kaplan-Meier", "Exponencial padrão"),
      lty = c(1, 3), lwd = 1, bty = "n", cex = 0.6)

# Gráfico 2: S(e) observado vs S(e) teórico
st <- ekm$surv
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1,
     xlab = "S(e): Kaplan-Meier", ylab = "S(e): Exponencial padrão",
     pch = 16)
abline(a = 0, b = 1, lty = 2, col = "black") # Linha de 45 graus

# Aproximação de Efron
Modelofinal_EFRON2<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = Testemunha2, ties = "efron")
summary(Modelofinal_EFRON2); anova(Modelofinal_EFRON2)
-2*as.numeric(logLik(Modelofinal_EFRON2))
#Verificando a pressuposição dos resíduo de schenfeld
resid(Modelofinal_EFRON2, type = "scaledsch")
par(mfrow=c(2,2))
cox.zph(Modelofinal_EFRON2, transform = "identity")
plot(cox.zph(Modelofinal_EFRON2))

Modelofinal_EXACT2<-coxph(Surv(time,stat)~APL+prec+temper, data = Testemunha2, ties = "exact")
summary(Modelofinal_EXACT2); anova(Modelofinal_EXACT2)

resid(Modelofinal_EXACT2, type = "scaledsch")
par(mfrow=c(2,2))
cox.zph(Modelofinal_EXACT2, transform = "identity")
plot(cox.zph(Modelofinal_EXACT2))

# Ajusta margens

```

```

par(mar = c(4, 4, 1, 2))

# Calcula resíduos de martingale
resm <- resid(Modelofinal_EXACT2, type = "martingale")
res <- Testemunha2$stat - resm

# Ajusta modelo de sobrevivência com os resíduos
ekm <- survfit(Surv(res, Testemunha2$stat) ~ 1)

# Define separador decimal como vírgula
options(OutDec = ",")

# Plota gráficos lado a lado
par(mfrow = c(1, 2))

# Gráfico 1: Curva Kaplan-Meier e curva exponencial teórica
plot(ekm, mark.time = FALSE, conf.int = FALSE,
     xlab = "Resíduos", ylab = "S(e) estimada")
res_ord <- sort(res)
exp1 <- exp(-res_ord)
lines(res_ord, exp1, lty = 3)
legend("topright", legend = c("Kaplan-Meier", "Exponencial padrão"),
      lty = c(1, 3), lwd = 1, bty = "n", cex = 0.6)

# Gráfico 2: S(e) observado vs S(e) teórico
st <- ekm$surv
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1,
     xlab = "S(e): Kaplan-Meier", ylab = "S(e): Exponencial padrão",
     pch = 16)
abline(a = 0, b = 1, lty = 2, col = "black") # Linha de 45 graus

# Dados agrupados
rm(list=ls())
ls()

library(readxl)
intervalos_t2 <- read_excel("Mestrado/DISSERTAÇÃO/Base de dados/intervalos_T2.xlsx")
head(intervalos_t2);attach(intervalos_t2)

# Modelo de riscos proporcionais de cox
modelo_tres <- glm(formula = y ~ int1 + Temp + int4 + int2 + APL + int5 +
                  Vig + NNR + Precip + factor(Fil), family = binomial(link = "cloglog"),
                  data = intervalos_t2)

summary(modelo_tres); anova(modelo_tres)
AIC(modelo_tres); BIC(modelo_tres)

# Curva ROC
{
  require(ROCR)

  # Ajustar separador decimal global
  options(OutDec = ",")

  # Predição e performance
  pred <- prediction(modelo_tres$fitted.values, intervalos_t2$y)
  perf <- performance(pred, "tpr", "fpr")
  auc <- performance(pred, "auc")@y.values[[1]]

  #par(mar = c(inferior, esquerda, superior, direita))
  par(mar = c(4, 4, 1, 1))

  # Plot vazio (sem eixos) para customização
  plot(perf,
       col = "blue",
       lwd = 2,

```

```

    main = "",
    xlab = "Taxa de falsos positivos",
    ylab = "Taxa de verdadeiros positivos",
    cex.lab = 1.2,
    cex.main = 1.4,
    xaxt = "n", # remove eixo X padrão
    yaxt = "n") # remove eixo Y padrão

# Customizar eixos com separador vírgula
axis(1, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))
axis(2, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))

# Linha diagonal de referência
abline(a = 0, b = 1, lty = 2, col = "black")

# Grid
grid()

# Exibir AUC no gráfico com vírgula
legend("bottomright",
      legend = paste("AUC =", format(round(auc, 4), decimal.mark = ",")),
      bty = "n")

# Exibir AUC no console
auc
}

# Modelo logístico
modelo_quatro <- glm(formula = y ~ int1 + Temp + int4 + int2 + APL + int5 +
                    Vig + NNR + Precip + factor(Fil), family = binomial(link = "logit"),
                    data = intervalos_t2)
summary(modelo_quatro);anova(modelo_quatro)
AIC(modelo_quatro);BIC(modelo_quatro)

# Curva ROC
{
  require(ROCR)

  # Ajustar separador decimal global
  options(OutDec = ",")

  # Predição e performance
  pred <- prediction(modelo_quatro$fitted.values, y)
  perf <- performance(pred, "tpr", "fpr")
  auc <- performance(pred, "auc")@y.values[[1]]

  #par(mar = c(inferior, esquerda, superior, direita))
  par(mar = c(4, 4, 3, 2))

  # Plot vazio (sem eixos) para customização
  plot(perf,
       col = "blue",
       lwd = 2,
       main = "Curva ROC",
       xlab = "Taxa de falsos positivos (FPR)",
       ylab = "Taxa de verdadeiros positivos (TPR)",
       cex.lab = 1.2,
       cex.main = 1.4,
       xaxt = "n", # remove eixo X padrão
       yaxt = "n") # remove eixo Y padrão

  # Customizar eixos com separador vírgula
  axis(1, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))
  axis(2, at = seq(0, 1, by = 0.1), labels = format(seq(0, 1, by = 0.1), decimal.mark = ","))

  # Linha diagonal de referência
  abline(a = 0, b = 1, lty = 2, col = "black")

```

```
# Grid
grid()

# Exibir AUC no gráfico com vírgula
legend("bottomright",
      legend = paste("AUC =", format(round(auc, 3), decimal.mark = ",")),
      bty = "n")

# Exibir AUC no console
auc
}
```

APÊNDICE C: Rotina para seleção de variáveis. O processo utilizado para as demais aproximações segue a mesma rotina.

```
# Seleção de covariáveis usando o modelo de regressão de Cox, com aproximação Breslow

#####
#####      Metodologia de Collet para aproximações      #####
#####      Aproximação de breslow                        #####
#####

# Carregando os dados

library(readxl)
Testemunha2 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/Testemunha_2.xlsx")
attach(Testemunha2); tail(Testemunha2)

library(survival)
covariaveis<-c("APL", "B1", "BM", "CER", "CRP", "DCA", "DCO", "FER", "Fil",
              "NNR", "NRP","Vig")

cafe<-Testemunha2
# Passo 1

{
  modelo_nulo<-coxph(Surv(time,stat)~1, data = cafe, ties = "breslow");summary(modelo_nulo)
  a1<--2*as.numeric(logLik(modelo_nulo));a1

  mod_APL<-coxph(Surv(time,stat)~APL, data = cafe, ties = "breslow");summary(mod_APL)
  ano1<-anova(modelo_nulo, mod_APL, test = "TRV");ano1
  a2<--2*as.numeric(logLik(mod_APL));a2

  mod_BLO<-coxph(Surv(time,stat)~B1, data = cafe, ties = "breslow");summary(mod_BLO)
  ano2<-anova(modelo_nulo,mod_BLO, test = "TRV");ano2
  a3<--2*as.numeric(logLik(mod_BLO));a3

  mod_BM<-coxph(Surv(time,stat)~BM, data = cafe, ties = "breslow");summary(mod_BM)
  ano3<-anova(modelo_nulo, mod_BM, test = "TRV");ano3
  a4<--2*as.numeric(logLik(mod_BM));a4

  mod_CER<-coxph(Surv(time,stat)~CER, data = cafe, ties = "breslow");summary(mod_CER)
  ano4<-anova(modelo_nulo, mod_CER, test = "TRV");ano4
  a5<--2*as.numeric(logLik(mod_CER));a5

  mod_CRP<-coxph(Surv(time,stat)~CRP, data = cafe, ties = "breslow");summary(mod_CRP)
  ano5<-anova(modelo_nulo, mod_CRP, test = "TRV");ano5
  a6<--2*as.numeric(logLik(mod_CRP));a6

  mod_DCA<-coxph(Surv(time,stat)~DCA, data = cafe, ties = "breslow");summary(mod_DCA)
  ano6<-anova(modelo_nulo, mod_DCA, test = "TRV");ano6
  a7<--2*as.numeric(logLik(mod_DCA));a7

  mod_DCO<-coxph(Surv(time,stat)~DCO, data = cafe, ties = "breslow");summary(mod_DCO)
  ano7<-anova(modelo_nulo, mod_DCO, test = "TRV");ano7
  a8<--2*as.numeric(logLik(mod_DCO));a8

  mod_FER<-coxph(Surv(time,stat)~FER, data = cafe, ties = "breslow");summary(mod_FER)
  ano8<-anova(modelo_nulo, mod_FER, test = "TRV");ano8
  a9<--2*as.numeric(logLik(mod_FER));a9

  mod_FIL<-coxph(Surv(time,stat)~Fil, data = cafe, ties = "breslow");summary(mod_FIL)
  ano9<-anova(modelo_nulo, mod_FIL, test = "TRV");ano9
  a10<--2*as.numeric(logLik(mod_FIL));a10

  mod_NNR<-coxph(Surv(time,stat)~NNR, data = cafe, ties = "breslow");summary(mod_NNR)
  ano10<-anova(modelo_nulo, mod_NNR, test = "TRV");ano10
}
```

```

a11<--2*as.numeric(logLik(mod_NNR));a11

mod_NRP<-coxph(Surv(time,stat)~NRP, data = cafe, ties = "breslow");summary(mod_NRP)
ano11<-anova(modelo_nulo, mod_NRP, test = "TRV");ano11
a12<--2*as.numeric(logLik(mod_NNR));a12

mod_VIG<-coxph(Surv(time,stat)~Vig, data = cafe, ties = "breslow");summary(mod_VIG)
ano12<-anova(modelo_nulo, mod_VIG, test = "TRV");ano12
a13<--2*as.numeric(logLik(mod_VIG));a13

mod_TEMP<-coxph(Surv(time,stat)~temper, data = cafe, ties = "breslow");summary(mod_TEMP)
ano13<-anova(modelo_nulo, mod_TEMP, test = "TRV");ano13
a14<--2*as.numeric(logLik(mod_TEMP));a14

mod_PREC<-coxph(Surv(time,stat)~prec, data = cafe, ties = "breslow");summary(mod_PREC)
ano14<-anova(modelo_nulo, mod_PREC, test = "TRV");ano14
a15<--2*as.numeric(logLik(mod_PREC));a15

Modelo1<-c("NULO", "APL", "Bl", "BM", "CER", "CRP", "DCA", "DCO", "FER", "Fil",
           "NNR", "NRP", "Vig", "temper", "prec")
logl<-c(a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11, a12, a13, a14, a15)
TRV1<-c("-", ano1$Chisq[2], ano2$Chisq[2], ano3$Chisq[2], ano4$Chisq[2], ano5$Chisq[2],
        ano6$Chisq[2], ano7$Chisq[2], ano8$Chisq[2], ano9$Chisq[2], ano10$Chisq[2],
        ano11$Chisq[2], ano12$Chisq[2], ano13$Chisq[2], ano14$Chisq[2])
P_valor1<-c("-", ano1$`Pr(>|Chi|)`[2], ano2$`Pr(>|Chi|)`[2], ano3$`Pr(>|Chi|)`[2], ano4$`Pr(>|Chi|)`[2],
            ano5$`Pr(>|Chi|)`[2], ano6$`Pr(>|Chi|)`[2], ano7$`Pr(>|Chi|)`[2], ano8$`Pr(>|Chi|)`[2], ano9$`Pr(>|Chi|)`[2],
            ano10$`Pr(>|Chi|)`[2], ano11$`Pr(>|Chi|)`[2], ano12$`Pr(>|Chi|)`[2], ano13$`Pr(>|Chi|)`[2],
            ano14$`Pr(>|Chi|)`[2])

Passo_1<-c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
Passo1<-cbind(Passo_1,Modelo1,logl,TRV1,P_valor1)
Passo1
}

#Passo 2
{
  modelo_atual<-coxph(Surv(time,stat)~APL+Bl+BM+DCA+FER+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow")
  atual<--2*as.numeric(logLik(modelo_atual));atual
  ano_atual<-anova(modelo_atual);ano_atual
  summary(modelo_atual)

  mods_APL<-coxph(Surv(time,stat)~Bl+BM+DCA+FER+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_APL)
  m_APL<--2*as.numeric(logLik(mods_APL));m_APL
  ano_APL<-anova(mods_APL, modelo_atual, test = "TRV");ano_APL

  mods_Bl<-coxph(Surv(time,stat)~APL+BM+DCA+FER+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_Bl)
  m_Bl<--2*as.numeric(logLik(mods_Bl));m_Bl
  ano_Bl<-anova(mods_Bl, modelo_atual, test = "TRV");ano_Bl

  mods_BM<-coxph(Surv(time,stat)~ APL+Bl+DCA+FER+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_BM)
  m_BM<--2*as.numeric(logLik(mods_BM));m_BM
  ano_BM<-anova(mods_BM, modelo_atual, test = "TRV");ano_BM

  mods_DCA<-coxph(Surv(time,stat)~APL+Bl+BM+FER+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_DCA)
  m_DCA<--2*as.numeric(logLik(mods_DCA));m_DCA
  ano_DCA<-anova(mods_DCA, modelo_atual, test = "TRV");ano_DCA

  mods_FER<-coxph(Surv(time,stat)~APL+Bl+BM+DCA+Fil+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_FER)
  m_FER<--2*as.numeric(logLik(mods_FER));m_FER
  ano_FER<-anova(mods_FER, modelo_atual, test = "TRV");ano_FER

  mods_FIL<-coxph(Surv(time,stat)~APL+Bl+BM+DCA+FER+NNR+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_FIL)
  m_FIL<--2*as.numeric(logLik(mods_FIL));m_FIL
  ano_FIL<-anova(mods_FIL, modelo_atual, test = "TRV");ano_FIL

  mods_NNR<-coxph(Surv(time,stat)~APL+Bl+BM+DCA+FER+Fil+Vig+temper+prec, data = cafe, ties = "breslow");summary(mods_NNR)
  m_NNR<--2*as.numeric(logLik(mods_NNR));m_NNR
  ano_NNR<-anova(mods_NNR, modelo_atual, test = "TRV");ano_NNR

```

```

mods_VIG<-coxph(Surv(time,stat)~APL+B1+BM+DCA+FER+Fil+NNR+temper+prec, data = cafe, ties = "breslow");summary(mods_VIG)
m_VIG<--2*as.numeric(logLik(mods_VIG));m_VIG
ano_VIG<-anova(mods_VIG, modelo_atual, test = "TRV");ano_VIG

mods_TEMP<-coxph(Surv(time,stat)~APL+B1+BM+DCA+FER+Fil+NNR+Vig+prec, data = cafe, ties = "breslow");summary(mods_TEMP)
m_TEMP<--2*as.numeric(logLik(mods_TEMP));m_TEMP
ano_TEMP<-anova(mods_TEMP, modelo_atual, test = "TRV");ano_TEMP

mods_PREC<-coxph(Surv(time,stat)~APL+B1+BM+DCA+FER+Fil+NNR+Vig+temper, data = cafe, ties = "breslow");summary(mods_PREC)
m_PREC<--2*as.numeric(logLik(mods_PREC));m_PREC
ano_PREC<-anova(mods_PREC, modelo_atual, test = "TRV");ano_PREC

Modelo2<-c("APL+B1+BM+DCA+FER+Fil+NNR+Vig+temper+prec", "B1+BM+DCA+FER+Fil+NNR+Vig+temper+prec", "APL+BM+DCA+FER+Fil+NNR+Vig+temper+prec",
"APL+B1+DCA+FER+Fil+NNR+Vig+temper+prec", "APL+B1+BM+FER+Fil+NNR+Vig+temper+prec",
"APL+B1+BM+DCA+Fil+NNR+Vig+temper+prec", "APL+B1+BM+DCA+FER+NNR+Vig+temper+prec",
"APL+B1+BM+DCA+FER+Fil+Vig+temper+prec", "APL+B1+BM+DCA+FER+Fil+NNR+temper+prec",
"APL+B1+BM+DCA+FER+Fil+NNR+Vig+prec", "APL+B1+BM+DCA+FER+Fil+NNR+Vig+temper")

looglike<-c(atual,m_APL,m_BL,m_BM,m_DCA,m_FER,m_FIL,m_NNR,m_VIG,m_TEMP,m_PREC)

TRV2<-c("-",ano_APL$Chisq[2],ano_BL$Chisq[2],ano_BM$Chisq[2],ano_DCA$Chisq[2],ano_FER$Chisq[2],ano_FIL$Chisq[2],
ano_NNR$Chisq[2],ano_VIG$Chisq[2],ano_TEMP$Chisq[2],ano_PREC$Chisq[2])

P_VALOR2<-c("-",ano_APL$`Pr(>|Chi|)`[2],ano_FIL$`Pr(>|Chi|)`[2],ano_BM$`Pr(>|Chi|)`[2],ano_DCA$`Pr(>|Chi|)`[2],
ano_FER$`Pr(>|Chi|)`[2],ano_FIL$`Pr(>|Chi|)`[2],ano_NNR$`Pr(>|Chi|)`[2],ano_VIG$`Pr(>|Chi|)`[2],ano_TEMP$`Pr(>|Chi|)`[2])

Passo_2<-c(2,2,2,2,2,2,2,2,2,2)
Passo2<-cbind(Passo_2,Modelo2,looglike,TRV2,P_VALOR2);Passo2
length(Passo_2)
length(Modelo2)
length(looglike)
length(TRV2)
length(P_VALOR2)
}
#APL+BM+DCA+FER+NNR+Vig+temper+prec
#Passo 3
{
modelo_P3<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = cafe, ties = "breslow")
atual3<--2*as.numeric(logLik(modelo_P3));atual3
ano_atual3<-anova(modelo_P3);ano_atual3
summary(modelo_P3)

mod_BL3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+B1, data = cafe, ties = "breslow");summary(mod_BL3)
m_BL3<--2*as.numeric(logLik(mod_BL3));m_BL3
ano_BL3<-anova(mod_BL3, modelo_P3, test = "TRV");ano_BL3

mod_BM3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+BM, data = cafe, ties = "breslow");summary(mod_BM3)
m_BM3<--2*as.numeric(logLik(mod_BM3));m_BM3
ano_BM3<-anova(mod_BM3, modelo_P3, test = "TRV");ano_BM3

mod_DCA3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+DCA, data = cafe, ties = "breslow");summary(mod_DCA3)
m_DCA3<--2*as.numeric(logLik(mod_DCA3));m_DCA3
ano_DCA3<-anova(mod_DCA3, modelo_P3, test = "TRV");ano_DCA3

mod_FER3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+FER, data = cafe, ties = "breslow");summary(mod_FER3)
m_FER3<--2*as.numeric(logLik(mod_FER3));m_FER3
ano_FER3<-anova(mod_FER3, modelo_P3, test = "TRV");ano_FER3

mod_FIL3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+Fil, data = cafe, ties = "breslow");summary(mod_FIL3)
m_FIL3<--2*as.numeric(logLik(mod_FIL3));m_FIL3
ano_FIL3<-anova(mod_FIL3, modelo_P3, test = "TRV");ano_FIL3

mod_NNR3<-coxph(Surv(time,stat)~APL+Vig+prec+temper+NNR, data = cafe, ties = "breslow");summary(mod_NNR3)
m_NNR3<--2*as.numeric(logLik(mod_NNR3));m_NNR3
ano_NNR3<-anova(mod_NNR3, modelo_P3, test = "TRV");ano_NNR3

```

```

Passo_3<-c(3,3,3,3,3,3)
Modelos3<-c("APL+Vig+prec+temper", "APL+Vig+prec+temper+BLO",
            "APL+Vig+prec+temper+BM", "APL+Vig+prec+temper+DCA", "APL+Vig+prec+temper+FER",
            "APL+Vig+prec+temper+FIL", "APL+Vig+prec+temper+NNR")

looglike3<-c(atual3,m_BL3,m_BM3,m_DCA3,m_FER3,m_FIL3,m_NNR3)

TRV3<-c("-",ano_BL3$Chisq[2],ano_BM3$Chisq[2],ano_DCA3$Chisq[2],ano_FER3$Chisq[2],
         ano_FIL3$Chisq[2], ano_NNR3$Chisq[2])

P_VALOR3<-c("-",ano_BL3$`Pr(>|Chi|)`[2], ano_BM3$`Pr(>|Chi|)`[2],
            ano_DCA3$`Pr(>|Chi|)`[2], ano_FER3$`Pr(>|Chi|)`[2],
            ano_FIL3$`Pr(>|Chi|)`[2], ano_NNR3$`Pr(>|Chi|)`[2])

passo3<-cbind(Passo_3,Modelos3,looglike3,TRV3,P_VALOR3);passo3
}

# Passo 4
{
Passo1
covariaveis<-c("APL", "B1", "BM", "CER", "CRP", "DCA", "DCO", "FER", "Fil",
              "NNR", "NRP", "Vig")

mod_p4<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = cafe, ties = "breslow"); summary(mod_p4)
atual4<--2*as.numeric(logLik(mod_p4)); atual4

mod_CER<-coxph(Surv(time,stat)~APL+Vig+prec+temper+CER, data = cafe, ties = "breslow"); summary(mod_CER)
m_CER<--2*as.numeric(logLik(mod_CER)); m_CER
ano_CER<-anova(mod_CER, mod_p4, test = "TRV");ano_CER

mod_CRP<-coxph(Surv(time,stat)~APL+Vig+prec+temper+CRP, data = cafe, ties = "breslow"); summary(mod_CRP)
m_CRP<--2*as.numeric(logLik(mod_CRP)); m_CRP
ano_CRP<-anova(mod_CRP, mod_p4, test = "TRV");ano_CRP

mod_DCO<-coxph(Surv(time,stat)~APL+Vig+prec+temper+DCO, data = cafe, ties = "breslow"); summary(mod_DCO)
m_DCO<--2*as.numeric(logLik(mod_DCO)); m_DCO
ano_DCO<-anova(mod_DCO, mod_p4, test = "TRV");ano_DCO

mod_NNR<-coxph(Surv(time,stat)~APL+Vig+prec+temper+NNR, data = cafe, ties = "breslow"); summary(mod_NNR)
m_NNR<--2*as.numeric(logLik(mod_NNR)); m_NNR
ano_NNR<-anova(mod_NNR, mod_p4, test = "TRV");ano_NNR

mod_NRP<-coxph(Surv(time,stat)~APL+Vig+prec+temper+NRP, data = cafe, ties = "breslow"); summary(mod_NRP)
m_NRP<--2*as.numeric(logLik(mod_NRP)); m_NRP
ano_NRP<-anova(mod_NRP, mod_p4, test = "TRV");ano_NRP

Passo_4<-c(4,4,4,4,4,4)
Modelos4<-c("APL+Vig+prec+temper",
            "APL+Vig+prec+temper+CER", "APL+Vig+prec+temper+CRP",
            "APL+Vig+prec+temper+DCO", "APL+Vig+prec+temper+NNR", "APL+Vig+prec+temper+NRP")

looglike4<-c(atual4,m_CER,m_CRP,m_DCO,m_FIL,m_NRP)

TRV4<-c("-",ano_CER$Chisq[2],ano_CRP$Chisq[2],ano_DCO$Chisq[2],ano_NNR$Chisq[2],
         ano_NRP$Chisq[2])

P_VALOR4<-c("-",ano_CER$`Pr(>|Chi|)`[2],ano_CRP$`Pr(>|Chi|)`[2],ano_DCO$`Pr(>|Chi|)`[2],ano_NNR$`Pr(>|Chi|)`[2],
            ano_NRP$`Pr(>|Chi|)`[2])

passo4<-cbind(Passo_4,Modelos4,looglike4,TRV4,P_VALOR4);passo4
}

# Passo 5
{
mod_p5<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = cafe, ties = "breslow"); summary(mod_p5)
atual5<--2*as.numeric(logLik(mod_p5)); atual5

mod_APL<-coxph(Surv(time,stat)~APL+prec+temper, data = cafe, ties = "breslow"); summary(mod_APL)

```

```

m_APL<--2*as.numeric(logLik(mod_APL)); m_APL
ano_APL<-anova(mod_APL, mod_p5, test = "TRV");ano_APL

mod_VIG<-coxph(Surv(time,stat)~APL+DCA, data = cafe, ties = "breslow"); summary(mod_VIG)
m_VIG<--2*as.numeric(logLik(mod_VIG)); m_VIG
ano_VIG<-anova(mod_VIG, mod_p5, test = "TRV");ano_VIG

mod_TEMP<-coxph(Surv(time,stat)~APL+Vig+prec, data = cafe, ties = "breslow"); summary(mod_TEMP)
m_TEMP<--2*as.numeric(logLik(mod_TEMP)); m_TEMP
ano_TEMP<-anova(mod_TEMP, mod_p5, test = "TRV");ano_TEMP

mod_PREC<-coxph(Surv(time,stat)~APL+Vig+temper, data = cafe, ties = "breslow"); summary(mod_PREC)
m_PREC<--2*as.numeric(logLik(mod_PREC)); m_PREC
ano_PREC<-anova(mod_PREC, mod_p5, test = "TRV");ano_PREC

Modelos5<-c("APL+Vig+prec+temper", "Vig+prec+temper",
            "APL+prec+temper", "APL+Vig+prec", "APL+Vig+temper")

looglike5<-c(atual5,m_APL,m_VIG,m_TEMP,m_PREC)

TRV5<-c("-",ano_APL$Chisq[2],ano_VIG$Chisq[2],ano_TEMP$Chisq[2], ano_PREC$Chisq[2])

P_VALOR5<-c("-",ano_APL$`Pr(>|Chi|)`[2],ano_VIG$`Pr(>|Chi|)`[2],
            ano_TEMP$`Pr(>|Chi|)`[2], ano_PREC$`Pr(>|Chi|)`[2])

Passo_5<-c(5,5,5,5,5)
passo5<-cbind(Passo_5,Modelos5,looglike5,TRV5,P_VALOR5);passo5
}

# Ajuste final
{
PASSO<-c(Passo_1,Passo_2,Passo_3,Passo_4,Passo_5);PASSO
MODELOS<-c(Modelo1,Modelo2,Modelos3,Modelos4,Modelos5);MODELOS
TRV<-c(TRV1,TRV2,TRV3,TRV4,TRV5);TRV
Looglike<-c(logl,looglike,looglike3,looglike4,looglike5);Looglike
P_VALOR<-c(P_valor1,P_VALOR2,P_VALOR3,P_VALOR4,P_VALOR5);P_VALOR

length(PASSO)
length(MODELOS)
length(TRV)
length(P_VALOR)
length(Looglike)
}

COLLETT_BRESLOW<-cbind(PASSO,MODELOS,Looglike,TRV,P_VALOR);COLLETT_BRESLOW

Modelofinal_BRESLOW<-coxph(Surv(time,stat)~APL+Vig+prec+temper, data = cafe, ties = "breslow")
summary(Modelofinal_BRESLOW); anova(Modelofinal_BRESLOW)
mf<--2*as.numeric(logLik(Modelofinal_BRESLOW));mf
#####
#####          Processo de seleção stepwise          #####
#####

#INTERVALO UM
rm(list=ls())
ls()
library(survival)
library(survminer)
library(readxl)
library(MASS)

intervalos_t1 <- read_excel("~/Mestrado/DISSERTAÇÃO/Base de dados/intervalos_T1.xlsx")
head(intervalos_t1)
attach(intervalos_t1)

# inicialmente para o modelo logit
names(intervalos_t1)

```

```

# Definir as covariáveis
covariaveis <- c("int1", "int2", "int3", "int4", "int5", "APL", "factor(BM)",
                "factor(CER)", "factor(B1)", "factor(Fil)", "CRP", "DCA",
                "NNR", "DCO", "factor(FER)", "NRP", "Precip", "Temp", "Vig")

# Criar a fórmula completa
formula_completa <- as.formula(paste("y ~", paste(covariaveis, collapse = " + ")))

# Definir o modelo inicial apenas com o intercepto
fit1 <- glm(y ~ 1, family = binomial(link = "cloglog"), data = intervalos_t1)

# Aplicar o método Stepwise com critério AIC
modelo_stepwise <- stepAIC(fit1,
                          scope = list(lower = ~1, upper = formula_completa),
                          direction = "both",
                          trace = TRUE) # trace = TRUE mostra o processo passo a passo

# Exibir o resumo do modelo final selecionado
summary(modelo_stepwise)

summary(modelo_stepwise);anova(modelo_stepwise)
library(car)
vif(modelo_stepwise)

modelo_um <- glm(formula = y ~ Temp + int1 + int2 + APL + int4 + Vig +
                NNR, family = binomial(link = "cloglog"), data = intervalos_t1)
summary(modelo_um);anova(modelo_um)
vif(modelo_um)
AIC(modelo_um)
BIC(modelo_um)
AIC(modelo_dois)
BIC(modelo_dois)

# Modelo logístico
# Definir as covariáveis adicionais que serão submetidas ao processo de Stepwise
# Carregar pacote necessário
library(MASS)

# Definir as covariáveis
covariaveis <- c("int1", "int2", "int3", "int4", "int5", "APL", "factor(BM)",
                "factor(CER)", "factor(B1)", "factor(Fil)", "CRP", "DCA",
                "NNR", "DCO", "factor(FER)", "NRP", "Precip", "Temp", "Vig")

# Criar a fórmula completa
formula_completa <- as.formula(paste("y ~", paste(covariaveis, collapse = " + ")))

# Definir o modelo inicial apenas com o intercepto
fit1 <- glm(y ~ 1, family = binomial(link = "logit"), data = intervalos_t1)

# Aplicar o método Stepwise com critério AIC
modelo_logit <- stepAIC(fit1,
                      scope = list(lower = ~1, upper = formula_completa),
                      direction = "both",
                      trace = TRUE) # trace = TRUE mostra o processo passo a passo

# Exibir o resumo do modelo final selecionado
summary(modelo_logit);anova(modelo_logit)

modelo_dois <- glm(formula = y ~ Temp + int1 + int2 + APL + int4 + Vig +
                NNR, family = binomial(link = "logit"), data = intervalos_t1)
AIC(modelo_dois)
BIC(modelo_dois)
summary(modelo_dois);anova(modelo_dois)

#INTERVALO DOIS
rm(list=ls())
ls()

```

```

library(readxl)
intervalos_t2 <- read_excel("Mestrado/DISSERTAÇÃO/Base de dados/intervalos_t2.xlsx")
head(intervalos_t2)
attach(intervalos_t2)
# Agora irei realizar o mesmo processo, entretanto considerando modelos discretos
# inicialmente para o modelo logit
names(intervalos_t2)
library(MASS)

# Definir as covariáveis
covariaveis <- c("int1", "int2", "int3", "int4", "int5", "APL", "factor(BM)",
                "factor(CER)", "factor(B1)", "factor(Fil)", "CRP", "DCA",
                "NNR", "DCO", "factor(FER)", "NRP", "Precip", "Temp", "Vig")

# Criar a fórmula completa
formula_completa <- as.formula(paste("y ~", paste(covariaveis, collapse = " + ")))

# Definir o modelo inicial apenas com o intercepto
fit1 <- glm(y ~ 1, family = binomial(link = "cloglog"), data = intervalos_t2)

# Aplicar o método Stepwise com critério AIC
modelo_stepwise <- stepAIC(fit1,
                          scope = list(lower = ~1, upper = formula_completa),
                          direction = "both",
                          trace = TRUE) # trace = TRUE mostra o processo passo a passo

# Exibir o resumo do modelo final selecionado
summary(modelo_stepwise)

modelo_tres <- glm(formula = y ~ int1 + Temp + int4 + int2 + APL + int5 +
                  Vig + NNR + Precip + factor(Fil), family = binomial(link = "cloglog"),
                  data = intervalos_t2)

summary(modelo_tres) # Exibe o resumo do modelo
anova(modelo_tres) # Exibe a análise de deviance
AIC(modelo_tres);BIC(modelo_tres)

# Modelo logístico

# Definir as covariáveis adicionais que serão submetidas ao processo de Stepwise
covariaveis <- c("int1","int2","int3","int4","int5","APL","factor(BM)"
                ,"factor(CER)","CRP","DCA","NNR","DCO","factor(FER)",
                "NRP","Precip","Temp","Vig")

# Definir o modelo inicial com as variáveis de intervalo fixas
fit1 <- glm(y ~ -1, family = binomial(link = "cauchit"), data = intervalos_t2)
# Criar a fórmula completa, incluindo as variáveis de intervalo fixas e as covariáveis adicionais

formula_completa <- as.formula(paste("y ~ -1+",
                                     paste(covariaveis, collapse = " + ")))

# Aplicar o método Stepwise com critério AIC
modelo_logit <- stepAIC(fit1, direction = "both", scope = list(lower = ~ -1,
                                                            upper = formula_completa))

# Exibir o resumo do modelo final selecionado
summary(modelo_logit);anova(modelo_logit)

modelo_quatro <- glm(formula = y ~ int1 + int4 + int3 + factor(CER) + Temp + int5 +
                    APL + Precip + Vig + NRP + NNR + CRP - 1, family = binomial(link = "logit"),
                    data = intervalos_t2)

# Suponha que 'intervalos_t2' é o conjunto de dados com as variáveis mencionadas
# Ajustar o modelo de regressão logística
modelo_quatro <- glm(formula = y ~ int1 + Temp + int4 + int2 + APL + int5 +
                    Vig + NNR + Precip + factor(Fil), family = binomial(link = "logit"),
                    data = intervalos_t2)
summary(modelo_quatro);anova(modelo_quatro)
AIC(modelo_quatro);BIC(modelo_quatro)

```