



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Tácio Cavalcante Xavier

Modelos de fração de cura log-simétricos com diferentes esquemas de ativação

Recife

2025

Tácio Cavalcante Xavier

Modelos de fração de cura log-simétricos com diferentes esquemas de ativação

Dissertação apresentada ao Programa de Pós-graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Estatística.

Área de Concentração: Estatística aplicada

Orientador (a): Francisco José de Azevêdo Cysneiros

Coorientador (a): Michelli Karinne Barros da Silva

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Xavier, Tácio Cavalcante.

Modelos de fração de cura log-simétricos com diferentes esquemas de ativação / Tácio Cavalcante Xavier. - Recife, 2025. 125f.: il.

Dissertação (Mestrado)- Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Programa de Pós-Graduação em Estatística, 2025.

Orientação: Francisco José de Azevêdo Cysneiros.

Coorientação: Michelli Karinne Barros da Silva.

1. Análise de sobrevivência; 2. Esquemas de ativação; 3. Fração de cura; 4. Modelos log-simétricos; 5. Resíduos de Cox-Snell. I. Cysneiros, Francisco José de Azevêdo. II. Silva, Michelli Karinne Barros da. III. Título.

UFPE-Biblioteca Central

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me conceder força e perseverança ao longo da vida.

À minha família, em especial à minha mãe, que sempre cuidou de mim com amor incondicional, enfrentando todas as dificuldades e me ensinando pelo exemplo.

Aos colegas e amigos do departamento: Maria, José, Francisco, Jhon, Yessenia, Joed e Eduarda, pela amizade e conversas que aliviaram a rotina com boas risadas.

Ao meu amigo Inácio que dividiu a moradia quando nos mudamos para o Recife, tornando mais leve minha primeira experiência longe de casa.

Aos professores Francisco José de Azevêdo Cysneiros e Michelli Karinne Barros da Silva, pela valiosa orientação e coorientação, pelo conhecimento generosamente compartilhado em cada reunião, pela paciência diante das minhas dúvidas e pela constante disponibilidade em me guiar.

À banca examinadora, pelo tempo dedicado, pelas sugestões e contribuições valiosas para o aprimoramento deste trabalho.

Aos demais professores do departamento que contribuíram significativamente para minha formação acadêmica.

À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pelo apoio financeiro.

A todos que, de alguma forma, caminharam ao meu lado durante este percurso, minha mais profunda gratidão.

RESUMO

Modelos de fração de cura são amplamente utilizados para analisar dados de tempo até o evento em contextos nos quais parte dos indivíduos nunca experimenta o evento de interesse. Este trabalho tem como objetivo desenvolver o modelo de fração de cura log-simétrico sob diferentes esquemas de ativação que combina diferentes mecanismos latentes relacionados à ocorrência da cura generalizando o trabalho de Rocha, Medeiros e Valença (2022). Desenvolvemos um método de estimação dos parâmetros baseado na verossimilhança marginal avaliando o comportamento dos estimadores baseado no viés relativo, desvio padrão e raiz do erro quadrático médio via estudo de simulação de Monte Carlo. A classe das distribuições log-simétricas, que inclui modelos assimétricos, estritamente positivos e com diferentes comportamentos de cauda, como por exemplo a log-normal e a log-t-Student, foi utilizada para modelar os tempos de vida dos indivíduos suscetíveis ao evento. Um estudo de simulação para avaliar o desempenho do resíduo proposto via resíduo Cox-Snell global para identificação de má especificação foi realizado considerando diferentes especificações incorretas do modelo. Por fim, a modelagem proposta foi aplicada a dados reais em estudos clínicos.

Palavras-chave: Análise de sobrevivência; Esquemas de ativação; Fração de cura; Modelos log-simétricos; Resíduos de Cox-Snell.

ABSTRACT

Cure fraction models are widely used to analyze time-to-event data in contexts where part of the individuals never experience the event of interest. This work aims to develop the log-symmetric cure fraction model under different activation schemes that combine different latent mechanisms related to the occurrence of cure, generalizing the work of Rocha, Medeiros e Valença (2022). We developed a parameter estimation method based on the marginal likelihood, evaluating the behavior of the estimators based on relative bias, standard deviation, and root mean squared error through a Monte Carlo simulation study. The class of log-symmetric distributions, which includes asymmetric, strictly positive models with different tail behaviors, such as the log-normal and the log-t-Student, was used to model the survival times of individuals susceptible to the event. A simulation study to evaluate the performance of the proposed residual using the global Cox–Snell residual for identifying misspecification was carried out considering different incorrect specifications of the model. Finally, the proposed modeling was applied to real data from clinical studies.

Keywords: Activation schemes; Cox–Snell residuals; Cure fraction; Log-symmetric models; Survival analysis.

LISTA DE TABELAS

Tabela 1 – Função geradora para alguns casos particulares da classe log-simétrica.	32
Tabela 2 – Vetores fixados para β nas simulações dos esquemas de primeira ativação e ativação aleatória	48
Tabela 3 – Vetores fixados para β nas simulações do esquema de ativação geométrica	48
Tabela 4 – viés relativo absoluto (VRA), raiz quadrada do erro quadrático médio (REQM) e desvio padrão (DP) das estimativas dos parâmetros no esquema de primeira ativação com latência log-normal, sob diferentes níveis de censura e cura.	51
Tabela 5 – VRA, REQM e DP das estimativas dos parâmetros no esquema de primeira ativação com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.	52
Tabela 6 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação aleatória com latência log-normal, sob diferentes níveis de censura e cura.	54
Tabela 7 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação aleatória com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.	55
Tabela 8 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.	58
Tabela 9 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.	59
Tabela 10 – VRA, REQM e DP das estimativas dos parâmetros no esquema de primeira ativação com latência log-normal, sob diferentes níveis de censura e cura.	61
Tabela 11 – VRA, REQM e DP das estimativas dos parâmetros no esquema de primeira ativação com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.	62
Tabela 12 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação aleatória com latência log-normal, sob diferentes níveis de censura e cura.	64
Tabela 13 – VRA, REQM e DP das estimativas dos parâmetros no esquema de ativação aleatória com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.	65

Tabela 14 – VRA, REQMe DP das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.	68
Tabela 15 – VRA, REQMe DP das estimativas dos parâmetros no esquema de ativação geométrica com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.	69
Tabela 16 – Vetores fixados para os β nas simulações dos esquemas de primeira ativação, ativação aleatória e ativação geométrica	71
Tabela 17 – Proporção de censura por grupo de tratamento	96
Tabela 18 – Proporção de censura por sexo	97
Tabela 19 – Modelos ajustados com diferentes esquemas de ativação	102
Tabela 20 – Estimativas dos coeficientes do modelo em investigação	103
Tabela 21 – Estimativas dos coeficientes do modelo final	103
Tabela 22 – Proporção de censura por grupo de tratamento	107
Tabela 23 – Proporção de censura por sexo	107
Tabela 24 – Proporção de censura por categoria do nódulo	108
Tabela 25 – Proporção de censura por status de performance	109
Tabela 26 – Modelos ajustados com diferentes esquemas de ativação	115
Tabela 27 – Estimativas dos coeficientes do modelo em investigação	116
Tabela 28 – Estimativas dos coeficientes do modelo	117

LISTA DE ABREVIATURAS E SIGLAS

AIC	critérios de informação de Akaike
BIC	critérios de informação de Bayes
DP	desvio padrão
EKM	estimador de Kaplan-Meier
fc	fração de cura
FDA	função de distribuição acumulada
FDP	função densidade de probabilidade
FS	função de sobrevivência
FSP	função de sobrevivência populacional
LS	log-simétrica
pcNC	proporção de censura nos não curados
REQM	raiz quadrada do erro quadrático médio
TRV	teste da razão de verossimilhança
VRA	viés relativo absoluto

SUMÁRIO

1	INTRODUÇÃO	10
1.1	MOTIVAÇÃO	14
1.2	OBJETIVOS	15
2	MODELOS DE FRAÇÃO DE CURA	17
2.1	MODELO DE MISTURA	19
2.2	MODELO TEMPO DE PROMOÇÃO	20
2.3	ESQUEMA DE PRIMEIRA ATIVAÇÃO	22
2.4	ESQUEMA DE ATIVAÇÃO ALEATÓRIA	26
2.5	ESQUEMA DE ATIVAÇÃO GEOMÉTRICA	29
3	MODELO DE FRAÇÃO DE CURA LOG-SIMÉTRICO COM ES-	
	QUEMAS DE ATIVAÇÃO	31
3.1	DISTRIBUIÇÃO LOG-SIMÉTRICA	31
3.2	MODELOS DE FRAÇÃO DE CURA LOG-SIMÉTRICOS	33
3.3	INFERÊNCIA	34
3.3.1	Modelo de primeira ativação log-simétrico	40
3.3.2	Modelo de ativação aleatória log-simétrico	42
3.3.3	Modelo de ativação geométrica log-simétrico	44
4	ESTUDO DE SIMULAÇÃO	47
4.1	ESTIMAÇÃO	47
4.1.1	Algoritmo 1	48
4.1.2	Cenário 1	50
4.1.3	Cenário 2	60
4.1.4	Conclusão	70
4.2	RESÍDUOS DE COX-SNELL	70
4.2.1	Resíduo de Cox-Snell	71
4.2.2	Resíduo de Cox-Snell com especificação correta	73
4.2.3	Resíduo de Cox-Snell com especificação incorreta na latência	79
4.2.4	Resíduo de Cox-Snell com especificação incorreta na incidência	82
4.2.5	Resíduo de Cox-Snell com especificação incorreta no esquema	88
4.2.6	Conclusão	94

5	APLICAÇÕES	96
5.1	APLICAÇÃO 1- ESTUDO CLÍNICO E1684	96
5.2	APLICAÇÃO 2- ESTUDO CLÍNICO E1690	106
6	CONSIDERAÇÕES FINAIS	120
	REFERÊNCIAS	122

1 INTRODUÇÃO

Em muitas áreas do conhecimento, surgem casos práticos no qual o interesse de uma pesquisa está no tempo até a ocorrência de um determinado evento (falha). A área da estatística que estuda esse tipo de caso é denominada análise de sobrevivência (ou confiabilidade). O evento de interesse pode variar conforme o contexto. Na engenharia, o evento de interesse geralmente se refere à ocorrência da falha de um determinado dispositivo, como, por exemplo, uma geladeira parar de funcionar. Já na área da saúde, o evento de interesse pode ser, por exemplo, a morte de um paciente, a detecção de um tumor cancerígeno, a interrupção do funcionamento de algum órgão, a remissão de uma doença ou a sua cura.

É crucial definir o evento de interesse com precisão para obter as informações desejadas. Por exemplo, no caso da geladeira, o evento pode ser definido como a paralisação completa do funcionamento ou como a redução, até certo ponto, da capacidade de preservar os alimentos, mesmo que ainda esteja operando. No caso do paciente, a falha pode ser a necessidade de iniciar uma diálise de emergência devido à falência renal aguda ou uma diálise decorrente de falência renal crônica, em que a perda da função renal é lenta, progressiva e irreversível. Nesse sentido, [Nelson \(2009\)](#) discute diferentes abordagens para a definição do evento de interesse no contexto de falhas.

Na análise de sobrevivência, a natureza dos dados desempenha um papel importante. Eles podem ser classificados como dados completos, quando o tempo exato até a ocorrência do evento de interesse é conhecido, ou como dados incompletos, quando há presença de censura. A censura ocorre quando se dispõe somente de informações parciais sobre o tempo até o evento, o que pode resultar de limitações orçamentárias, tempo insuficiente de acompanhamento, perdas durante o seguimento ou do próprio delineamento do experimento.

Na presença de censura, métodos estatísticos usuais de inferência que desconsideram o fato de que as observações são parciais não são recomendados, pois podem resultar em estimativas enviesadas e na perda de informação. Nesse sentido, o entendimento adequado da censura é essencial para que essa informação seja corretamente incorporada ao modelo estatístico. A censura pode ser classificada como censura por intervalo, quando não se sabe exatamente quando o evento de interesse ocorreu, mas sabe-se que ele ocorreu entre dois momentos específicos. Por exemplo, em exames médicos periódicos, um paciente pode não apresentar sinais de uma doença em um exame, mas ser diagnosticado na consulta seguinte. Nesse caso,

o tempo até contrair a doença está compreendido entre os dois exames, mas é desconhecido. Dois casos particulares de censura por intervalo são a censura à esquerda, que ocorre quando não se sabe exatamente o momento em que o evento de interesse aconteceu, mas sabe-se que ele ocorreu antes do tempo registrado no estudo, e a censura à direita, quando também não se conhece o momento exato do evento, mas sabe-se que ele ocorreu após o tempo observado. Além dessa categorização, a censura à direita pode ser subdividida em três tipos. Na censura do tipo I (ou censura por tempo), onde define-se um tempo máximo de observação, e todas as unidades que não apresentarem o evento até esse ponto são censuradas nesse tempo máximo; censura do tipo II (ou censura por falha), onde se estabelece, a priori, um número fixo de eventos a ser observado e ao atingir esse número, as demais unidades que não apresentaram o evento de interesse são censuradas no tempo em que a última observação apresentou o evento. Já na censura aleatória, o tempo até o evento de interesse é interrompido devido a fatores não controlados, impedindo a observação do evento. Por exemplo, em um estudo sobre o tempo até a morte de pacientes com câncer submetidos a um novo tratamento, um paciente que falece em um acidente de carro tem o tempo de observação censurado, pois sua morte ocorreu por uma causa não relacionada ao evento de interesse. Para mais detalhes sobre os tipos de censura, ver Colosimo e Giolo (2006) e Nelson (2009).

Algumas suposições sobre o mecanismo da censura são comumente adotadas, uma delas é o esquema de censura aleatória, na qual cada observação censurada na amostra tem taxa de risco igual à de uma observação não censurada. Uma suposição menos restritiva, que decorre da anterior é a de censura independente, em que, dentro de qualquer subgrupo de interesse da amostra, por exemplo, tratamento e controle, a censura é aleatória, ou seja, condicional ao subgrupo, a taxa de risco de uma observação censurada é igual à de uma observação não censurada. Por fim, considera-se a hipótese de censura não informativa, segundo a qual os tempos de censura e os tempos completos são independentes entre si, isto é, informações sobre um não fornecem qualquer conhecimento sobre o outro. Essas suposições são abordadas detalhadamente em Kleinbaum e Klein (1996) e Kalbfleisch e Prentice (2011).

Para além das suposições sobre o mecanismo da censura, é necessário também fazer suposições sobre a distribuição dos tempos das observações não censuradas. Em muitos modelos estatísticos, assume-se a normalidade da variável de interesse. No entanto, essa suposição é frequentemente inválida, devido à, por exemplo, assimetria dos dados ou fato de os dados serem estritamente positivos, ou negativos. Uma possível alternativa, nesse caso em que os dados são estritamente positivos e assimétricos, é utilizar a distribuição log-normal que é

amplamente conhecida na modelagem de dados assimétricos (JOBE, 1989).

Porém, a distribuição log-normal é apenas uma das possíveis alternativas e pode não ser adequada em situações, por exemplo, em que os dados apresentam bimodalidade. Nesse sentido, visando abranger um maior número de casos reais, a suposição de normalidade da variável de interesse, ou dos erros aleatórios, pode ser substituída pela suposição de que a distribuição pertence à classe de distribuições log-simétrica (LS). Essa classe é adequada para modelar dados estritamente positivos e assimétricos, podendo acomodar características como bimodalidade e caudas mais leves ou mais pesadas do que as da distribuição log-normal.

O primeiro trabalho com o intuito de caracterizar a classe de distribuições LS foi realizado por Seshadri (1965), que estudou algumas de suas propriedades e demonstrou que uma condição necessária e suficiente para a existência dessa classe é que a distribuição do seu logaritmo seja uma distribuição simétrica. Posteriormente, Jones e Arnold (2008) propuseram uma nova classe de distribuições que são simultaneamente log-simétricas e R-simétricas, denominada classe duplamente simétrica. Vanegas e Paula (2016) detalharam as propriedades estatísticas da classe LS, derivando medidas de interesse baseadas em quantis, bem como as expressões das funções geradoras de densidade e de sua derivada para diversas distribuições pertencentes à classe, além disso, também propuseram métodos de estimação tanto no contexto clássico quanto no bayesiano para várias dessas distribuições. Já Paiva (2020) discutiram uma versão discreta da classe log-simétrica, abordando suas propriedades e métodos de estimação clássica. Por sua vez, Saulo et al. (2022) apresentaram um modelo de regressão quantílica baseado em uma reparametrização da classe LS, denominada classe log-simétrica quantílica. Posteriormente, Cunha, Divino e Saulo (2022) e Cunha, Divino e Saulo (2024) utilizaram essa classe em modelos Tobit quantílico e de regressão quantílica zero-ajustada, respectivamente. Já Vila et al. (2023) estenderam a classe LS para o caso bivariado, apresentando suas propriedades e a estimação dos parâmetros sob a abordagem clássica.

A construção da classe de distribuições LS é realizada por meio da transformação exponencial de uma variável aleatória pertencente à classe das distribuições simétricas. Essa abordagem abrange diversas distribuições para dados positivos assimétricos como casos particulares, tais como: log-normal, log-t-Student, Harmonic law, Birnbaum-Saunders, Birnbaum-Saunders-t, Birnbaum-Saunders generalizada, log-normal contaminada, log-exponencial-potência, log-slash, entre outras (VANEGAS; PAULA, 2016).

O conceito de log-simetria dessa classe pode ser compreendido em termos da propriedade de fechamento sob reciprocidade, ou seja, uma variável aleatória que possui a mesma distri-

buição que sua recíproca, pertence a classe [LS](#). Essa característica é uma das propriedades fundamentais da classe, conforme demonstrado em [\(SESHADRI, 1965\)](#), [\(JONES, 2008\)](#) e [\(VANEGAS; PAULA, 2016\)](#). Outra propriedade importante é que um dos parâmetros da distribuição pode ser interpretado diretamente como a mediana, sendo, portanto, o parâmetro de locação da distribuição. Essa interpretação é particularmente útil, pois a mediana é uma medida de tendência central mais robusta do que a média na presença de *outliers*.

Um dos pressupostos básicos da análise de sobrevivência é que, dado tempo suficiente, todas as observações eventualmente apresentarão o evento de interesse. No entanto, esse pressuposto pode não ser válido em diversas situações. Por exemplo, se o evento de interesse for a recidiva de uma doença, um paciente curado terá um tempo até a recidiva infinito, o que, na prática, significa que o evento nunca será observado. Com o objetivo de contemplar cenários em que uma parcela da população jamais apresentará o evento de interesse, foram desenvolvidos, inicialmente na área da saúde, os modelos de fração de cura (ou modelos de longa duração). Posteriormente, essa abordagem foi aplicada em diferentes áreas, recebendo outras denominações: na engenharia, são conhecidos como modelos de vida com população de falha limitada [\(MEEKER, 1987\)](#); e, na economia, como modelos de população dividida [\(SCHMIDT; WITTE, 1989\)](#).

O primeiro modelo de fração de cura, conhecido como modelo de mistura, foi introduzido por [Boag \(1949\)](#). Nesse modelo, proposto no contexto da terapia para câncer de mama, a população é dividida entre curados e não curados, por meio de uma variável binária indicadora. Outro modelo amplamente utilizado é uma extensão do modelo de riscos proporcionais de cox [\(COX, 1972\)](#), adaptado para incorporar a fração de cura. Trata-se do modelo de tempo de promoção, proposto e desenvolvido por [Tsodikov, Yakovlev e Asselain \(1996\)](#) e [Chen, Ibrahim e Sinha \(1999\)](#). Neste modelo é incorporada uma estrutura hierárquica, na qual assume-se que uma observação poderá apresentar o evento de interesse se for exposta a pelo menos um fator latente causador do evento. O número desses fatores segue uma distribuição de Poisson. Caso a observação não esteja associada a nenhum fator, ela é considerada curada. Cada fator latente possui um tempo próprio até atingir um determinado limite e, então, causar o evento. Para as observações suscetíveis (ou seja, com pelo menos um fator), o tempo até o evento corresponde ao mínimo dos tempos latentes entre os fatores, isto é, ao primeiro fator a atingir o limite. Ao longo do tempo, diversas modificações desses dois modelos originais foram propostas. Por exemplo, [Li e Taylor \(2002\)](#) introduziram ao modelo de mistura uma estrutura semiparamétrica baseada no modelo de vida acelerada.

Devido às diferentes abordagens de modelagem desenvolvidas para os modelos de fração de cura, surgiu a necessidade de uma estrutura unificada que englobasse essas propostas. Uma das primeiras tentativas nesse sentido foi apresentada por Yin e Ibrahim (2005), que utilizou a transformação de Box-Cox aplicada à função de sobrevivência. Nesse modelo, os modelos de mistura e de tempo de promoção surgem como casos particulares, quando o parâmetro da transformação assume os valores um e zero, respectivamente. Outra abordagem proposta por Cooner et al. (2007) amplia e flexibiliza a estrutura hierárquica do modelo de tempo de promoção ao permitir que o número de fatores latentes siga qualquer distribuição discreta positiva e ao considerar que o evento de interesse ocorre somente quando um número mínimo desses fatores atinge um limite crítico menor ou igual ao número de fatores latentes. O tempo até o evento, nesse caso, corresponde ao momento em que o último desses fatores atinge esse limite. A depender do número mínimo de fatores necessários, obtêm-se diferentes modelos, denominados esquemas de ativação, sendo que essa quantidade pode ser uma constante, uma função da quantidade de fatores necessários ou mesmo uma distribuição aleatória. Quando esse número mínimo segue uma distribuição uniforme discreta, tem-se o esquema aleatório; quando é constante e igual a 1, tem-se o esquema de primeira ativação (RODRIGUES et al., 2009); e quando é igual ao número total de fatores latentes, tem-se o esquema de última ativação. Nesse contexto, os modelos de mistura e de tempo de promoção são obtidos como casos especiais, de acordo com a distribuição assumida para o número de fatores latentes no esquema de primeira ativação. Posteriormente, uma proposta ainda mais geral foi apresentada por Kim, Chen e Dey (2011), que flexibiliza o número mínimo de fatores latentes necessários, permitindo que esse número seja superior ao total de fatores latentes presentes, ampliando ainda mais a generalização do modelo.

1.1 MOTIVAÇÃO

Entre os diversos modelos desenvolvidos para lidar com dados de sobrevivência, os modelos de fração de cura ocupam lugar de destaque por permitirem incorporar a possibilidade de que uma parcela da população jamais experimente o evento de interesse, como pacientes efetivamente curados ou dispositivos imunes a falhas. Essa característica rompe com a suposição clássica de que, dado tempo suficiente, todos os indivíduos sofrerão eventualmente o evento. Assim, os modelos de fração de cura têm sido amplamente utilizados na prática e vêm sendo continuamente aprimorados para incorporar diferentes estruturas, distribuições e mecanismos

latentes.

Entretanto, apesar dos avanços na literatura, muitas propostas consideram a estrutura de número de fatores necessário para ocorrência do evento como sendo de primeira ativação. Isso pode limitar a flexibilidade dos modelos e consequentemente, sua capacidade de captar adequadamente a complexidade dos dados empíricos. Nesse contexto, torna-se necessária a busca por estruturas mais abrangentes e versáteis, capazes de englobar modelos existentes como casos particulares e de permitir a modelagem de diferentes comportamentos para o número de fatores necessários à ocorrência do evento.

Nesse cenário, esta dissertação propõe-se a trabalhar com um modelo mais abrangente de fração de cura, capaz de incorporar, como casos particulares, modelos clássicos como o de mistura e o de tempo de promoção, além de variantes mais recentes que consideram diferentes combinações entre o número de causas latentes e o número necessário para ocorrência do evento. Além da flexibilidade do modelo de fração de cura adotado, esta dissertação utiliza a classe das distribuições log-simétricas para modelar os tempos de vida dos indivíduos suscetíveis. Trata-se de uma classe abrangente, capaz de acomodar diferentes graus de assimetria, de caudas nos dados e que inclui, como casos particulares, distribuições amplamente utilizadas, como a log-normal, a log-t-Student, a Birnbaum-Saunders e suas variantes, entre outras.

1.2 OBJETIVOS

O principal objetivo deste trabalho é generalizar o modelo de fração de cura log-simétrico de Rocha, Medeiros e Valença (2022), no qual é utilizado a classe de distribuições log-simétricas para modelar os tempos de vida dos indivíduos suscetíveis, para um modelo que admite diferentes esquemas de ativação. Essa generalização permite uma quantidade maior de combinações entre o número de causas latentes e o número necessário para a ocorrência do evento.

Outro objetivo é o desenvolvimento de um extenso estudo de simulação para avaliar o comportamento dos estimadores de máxima verossimilhança utilizando verossimilhança marginal sob diferentes configurações de esquemas de ativação, como primeira ativação, ativação aleatória e ativação geométrica, considerando distribuições da classe log-simétrica, como a log-normal e a log-t-Student com três graus de liberdade. Esse estudo considerará diferentes cenários, variando a fração de cura, a proporção de censura e o tamanho amostral. O desempenho dos estimadores do modelo foi avaliado com base em métricas como a raiz quadrada do erro quadrático médio (REQM), o viés relativo absoluto (VRA) e a análise dos resíduos de

Cox–Snell, especialmente em situações com especificação incorreta da estrutura do modelo.

Para ilustrar a metodologia desenvolvida, analisamos dois conjuntos de dados reais, nos quais há evidências da presença de indivíduos curados, justificando o uso de modelos de fração de cura. Para a seleção do modelo mais adequado, foram utilizados o **critérios de informação de Akaike (AIC)**, **critérios de informação de Bayes (BIC)** e **teste da razão de verossimilhança (TRV)** para modelos encaixados. A avaliação da qualidade do ajuste foi feita por meio da análise dos resíduos de Cox–Snell e de comparações das curvas estimadas pelo modelo e pelo **estimador de Kaplan-Meier (EKM)**.

2 MODELOS DE FRAÇÃO DE CURA

Na análise de sobrevivência a atenção está em estudar o tempo até a ocorrência de um evento de interesse, por exemplo, o tempo até a inadimplência em contratos de crédito. Esse tempo pode ser analisado em termos probabilísticos por meio da **função de sobrevivência (FS)**.

Definição 1. Seja T uma variável aleatória contínua e não negativa representando o tempo até o evento de interesse. A função de sobrevivência, denotada por $S(t)$, é definida como a probabilidade de uma observação não apresentar o evento de interesse até o tempo t , ou seja, a probabilidade de uma observação sobreviver além do tempo t , dada por

$$S(t) = 1 - F(t),$$

em que $F(t) = P(T \leq t)$ é a da variável aleatória T .

Propriedades. A função de sobrevivência $S(t)$ possui as seguintes propriedades:

- I) $S(t)$ é contínua para todo t não-negativo;
- II) $S(0) = 1$;
- III) $\lim_{t \rightarrow +\infty} S(t) = 0$;
- IV) $S(t)$ é monótona não crescente, ou seja, $S(t) \leq S(t')$ para todo $t \geq t'$.

Em muitos casos, nem todas as observações de uma amostra são suscetíveis ao evento de interesse. Nesses cenários, a condição de normalização, uma das propriedades fundamentais da **função densidade de probabilidade (FDP)**, não é satisfeita e passa a ser dada por

$$\int_0^{\infty} f(u) du = \Upsilon \leq 1.$$

Nesse cenário, considera-se uma classe mais ampla de funções, denominada funções impróprias, da qual a **FDP** usual é um caso particular.

Definição 2. Dada uma amostra aleatória de uma variável aleatória T não negativa e contínua, a presença de um mecanismo latente que leva a variável de parte da amostra ao ∞ , a **função**



de sobrevivência populacional (FSP), denotada por, $S_{pop}(t)$, é dada por

$$\begin{aligned}
 S_{pop}(t) &= P(T > t) \\
 &= 1 - P(T \leq t) \\
 &= 1 - P(0 \leq T \leq t) \\
 &= 1 - P(T \leq t) + P(T \leq 0) \\
 &= 1 - 1 + P(T > t) + 1 - P(T > 0) \\
 &= 1 - P(T > 0) + P(T > t) \\
 &= 1 - \int_0^\infty f(u) du + \int_t^\infty f(u) du \\
 &= 1 - \Upsilon + S(t),
 \end{aligned}$$

em que $S(t)$ é a FS dos indivíduos suscetíveis ao evento de interesse na amostra.

Propriedades. A FSP $S_{pop}(t)$ possui as seguintes propriedades:

- I) $S_{pop}(t)$ é uma mistura da parte degenerada em $1 - \Upsilon$ e da parte contínua $S(t)$ para todo t não negativo;
- II) $S_{pop}(0) = 1$;
- III) $\lim_{t \rightarrow +\infty} S_{pop}(t) = 1 - \Upsilon$, onde $1 - \Upsilon$ é denominado fração de cura, com interpretação dependendo do contexto dos dados;
- IV) $S_{pop}(t)$ é monótona não crescente, ou seja, $S_{pop}(t) \leq S_{pop}(t')$ para todo $t \geq t'$;
- V) Se $\Upsilon = 1$, então, encontra-se a classe usual de funções de densidade, ou seja, $S_{pop}(t) = S(t)$.

Quando a Propriedade III da FS não é atendida, isto é, quando parte dos indivíduos da amostra não apresenta o evento de interesse, diz-se que há uma fração de cura na amostra. Algumas vezes, a existência dessa fração pode ser identificada por especialistas da área de estudo, com base em evidências já estabelecidas na literatura ou em critérios que definem quando um indivíduo pode ser considerado curado. No entanto, nem sempre essas informações estão disponíveis. Nesses casos, é necessário recorrer a métodos alternativos para identificar a presença da fração de cura. Maller e Zhou (1996) propõem que essa identificação seja feita a partir da curva de Kaplan–Meier (KAPLAN; MEIER, 1958), observando-se a formação de um platô na estimativa da FS pelo EKM. Esse platô indicaria que, a partir de certo ponto, a

probabilidade de sobrevivência se mantém constante, sugerindo a presença de indivíduos que não experimentarão o evento de interesse.

Contudo, é importante ressaltar que a presença de um platô pode ser consequência de um tempo de acompanhamento insuficiente. Nessa situação, indivíduos que ainda não apresentaram o evento podem fazê-lo futuramente. Por isso, a recomendação é considerar a existência de um platô longo na curva de sobrevivência estimada pelo estimador de Kaplan-Meier, sendo que a definição de “longo” depende do contexto dos dados ou do seu delineamento experimental.

2.1 MODELO DE MISTURA

O primeiro modelo a incorporar a fração de cura é o modelo de mistura proposto por Boag (1949) e desenvolvido em Berkson e Gage (1952), no qual se assume que a população é composta por dois grupos: curados e não curados. Para os indivíduos curados, o tempo até o evento de interesse é considerado infinito; já para os não curados, esse tempo segue uma distribuição contínua e não negativa. A divisão dos indivíduos para os dois grupos é modelada como uma variável aleatória com distribuição Bernoulli.

Seja M uma variável dicotômica definida no mesmo espaço amostral que a variável de tempo até o evento de interesse, $\tilde{Y}$, em que $M = 0$ indica que o indivíduo está sujeito ao evento, ou seja, possui um tempo finito $\tilde{Y}$, e $M = 1$ indica que o indivíduo está *curado*, ou seja, seu tempo até o evento é infinito. Supondo independência entre $\tilde{Y}$ e M , o tempo populacional até a ocorrência do evento pode ser escrito como

$$Y_{\text{pop}} = \infty \times I(M = 1) + \tilde{Y} \times I(M = 0).$$

Portanto, a função de sobrevivência populacional do modelo de mistura fica dada por

$$\begin{aligned} S_{\text{pop}}(y) &= P(Y_{\text{pop}} > y, M = 1) + P(Y_{\text{pop}} > y, M = 0) \\ &= P(Y_{\text{pop}} > y \mid M = 1)P(M = 1) + P(Y_{\text{pop}} > y \mid M = 0)P(M = 0) \\ &= P(M = 1) + P(\tilde{Y} > y)P(M = 0) \\ &= P(M = 1) + [1 - P(M = 1)]S_{\tilde{Y}}(y), \end{aligned}$$

em que $P(\tilde{Y} > y \mid M = 0) = P(\tilde{Y} > y)$ devido à independência entre $\tilde{Y}$ e M .

Analisando suas propriedades, temos

$$\begin{aligned} S_{\text{pop}}(0) &= P(M = 1) + [1 - P(M = 1)]S_{\tilde{Y}}(0) \\ &= P(M = 1) + [1 - P(M = 1)] \times 1 \\ &= 1, \end{aligned}$$

$$\begin{aligned} \lim_{y \rightarrow \infty} S_{\text{pop}}(y) &= P(M = 1) + [1 - P(M = 1)] \lim_{y \rightarrow \infty} S_{\tilde{Y}}(y) \\ &= P(M = 1) + [1 - P(M = 1)] \times 0 \\ &= P(M = 1) \\ &= p_0, \end{aligned}$$

em que $p_0 = P(M = 1)$ representa a fração de cura na população.

2.2 MODELO TEMPO DE PROMOÇÃO

Outro modelo famoso é o modelo tempo de promoção (TSODIKOV; YAKOVLEV; ASSELAIN, 1996) (CHEN; IBRAHIM; SINHA, 1999) que se diferencia do modelo de mistura por possuir a propriedade de riscos proporcionais e oferecer uma interpretação biológica. Esse modelo introduz uma estrutura hierárquica ao considerar o tempo até a falha sujeito à presença de fatores latentes. O tempo até o evento corresponde ao primeiro fator latente que atinge um limiar de ativação, sendo que a quantidade de fatores é uma variável aleatória com distribuição Poisson.

Considere que M denota o número de fatores latentes competindo para o evento de interesse, por exemplo, o número de células tumorais competindo para metástase que permanecem ativas após o tratamento inicial, e que $\tilde{Y}_i$ denote o tempo para o i -ésimo fator ocasionar o evento de interesse, por exemplo, o tempo para a i -ésima célula tumoral produzir uma doença metastática. Assim, $\tilde{Y}_i$ pode ser interpretado como o tempo de incubação necessário para que a i -ésima célula tumoral gere uma massa de câncer detectável. Suponha que $M \sim \text{Poisson}(\theta)$, que as variáveis $\tilde{Y}_i, i = 1, 2, \dots, m$, são independentes e identicamente distribuídas e que M e os $\tilde{Y}_i$ são independentes. O tempo até o evento de interesse, como por exemplo, a detecção do câncer, pode ser escrito como

$$Y_{\text{pop}} = \min(\tilde{Y}_i : 0 \leq i \leq M),$$

com $P(\tilde{Y}_0 = \infty) = 1$, ou equivalentemente,

$$Y_{\text{pop}} = \infty \times I(M = 0) + \tilde{Y}_{(1)} \times I(M \geq 1),$$

em que $\tilde{Y}_{(1)}$ é o menor dos $\tilde{Y}_i$.

Portanto, a **FSP** do modelo tempo de promoção é dada por

$$\begin{aligned}
 S_{Y_{\text{pop}}}(y) &= \sum_{m=0}^{\infty} P(Y_{\text{pop}} > y, M = m) \\
 &= \sum_{m=0}^{\infty} P(Y_{\text{pop}} > y \mid M = m)P(M = m) \\
 &= P(Y_{\text{pop}} > y \mid M = 0)P(M = 0) + \sum_{m=1}^{\infty} P(Y_{\text{pop}} > y \mid M = m)P(M = m) \\
 &= P(M = 0) + \sum_{m=1}^{\infty} P(\min(\tilde{Y}_1, \dots, \tilde{Y}_m) > y \mid M = m)P(M = m) \\
 &= P(M = 0) + \sum_{m=1}^{\infty} P(\tilde{Y}_1 > y, \tilde{Y}_2 > y, \dots, \tilde{Y}_m > y \mid M = m)P(M = m) \\
 &= P(M = 0) + \sum_{m=1}^{\infty} P(\tilde{Y}_1 > y, \tilde{Y}_2 > y, \dots, \tilde{Y}_m > y)P(M = m) \\
 &= \exp\{-\theta\} + \sum_{m=1}^{\infty} P(\tilde{Y}_1 > y)^m \frac{\exp\{-\theta\}\theta^m}{m!} \\
 &= \exp\{-\theta\} \left[1 + \sum_{m=1}^{\infty} S_{\tilde{Y}}(y)^m \frac{\theta^m}{m!} \right] \\
 &= \exp\{-\theta\} \left\{ \sum_{m=0}^{\infty} \frac{[\theta S_{\tilde{Y}}(y)]^m}{m!} \right\} \\
 &= \exp\{-\theta\} \exp\{\theta S_{\tilde{Y}}(y)\} \\
 &= \exp\{-\theta F_{\tilde{Y}}(y)\},
 \end{aligned}$$

em que $P(\tilde{Y}_1 > y, \tilde{Y}_2 > y, \dots, \tilde{Y}_m > y \mid M = m) = P(\tilde{Y}_1 > y, \tilde{Y}_2 > y, \dots, \tilde{Y}_m > y)$ devido à independência entre os $\tilde{Y}_i$'s e M .

Analisando suas propriedades, temos

$$\begin{aligned}
 S_{Y_{\text{pop}}}(0) &= \exp\{-\theta F_{\tilde{Y}}(0)\} \\
 &= \exp\{-\theta \times 0\} \\
 &= 1,
 \end{aligned}$$

$$\begin{aligned}
 \lim_{y \rightarrow \infty} S_{Y_{\text{pop}}}(y) &= \exp\{-\theta \times \lim_{y \rightarrow \infty} F_{\tilde{Y}}(y)\} \\
 &= \exp\{-\theta \times 1\} \\
 &= \exp\{-\theta\} \\
 &= p_0,
 \end{aligned}$$

em que $p_0 = \exp\{-\theta\}$ representa a fração de cura na população.

2.3 ESQUEMA DE PRIMEIRA ATIVAÇÃO

Rodrigues et al. (2009) em seu trabalho, retiraram a suposição de distribuição Poisson para M . Dessa forma, o tempo de vida ainda é escrita da mesma forma que no modelo tempo de promoção, porém sua função de sobrevivência é alterada.

Considere, assim como anteriormente, que M denote o número de fatores latentes competindo para o evento de interesse, e que $\tilde{Y}_i$ denote o tempo para o i -ésimo fator ocasionar o evento de interesse. Suponha que as variáveis $\tilde{Y}_i$, para $i = 1, 2, \dots, m$, sejam independentes e identicamente distribuídas, e que M e os $\tilde{Y}_i$ sejam independentes. O tempo até o evento de interesse pode ser escrito como

$$Y_{\text{pop}} = \infty \times I(M = 0) + \tilde{Y}_{(1)} \times I(M \geq 1). \quad (2.1)$$

Portanto, a função de sobrevivência populacional do modelo é dada por

$$\begin{aligned} S_{Y_{\text{pop}}}(y) &= \sum_{m=0}^{\infty} P(Y_{\text{pop}} > y, M = m) \\ &= \sum_{m=0}^{\infty} P(Y_{\text{pop}} > y | M = m) P(M = m) \\ &= P(Y_{\text{pop}} > y | M = 0) P(M = 0) + \sum_{m=1}^{\infty} P(Y_{\text{pop}} > y | M = m) P(M = m) \\ &= P(M = 0) + \sum_{m=1}^{\infty} P(\min(\tilde{Y}_1, \dots, \tilde{Y}_m) > y | M = m) P(M = m) \\ &= P(M = 0) + \sum_{m=1}^{\infty} S_{\tilde{Y}}(y)^m P(M = m). \end{aligned} \quad (2.2)$$

Analisando suas propriedades, temos

$$\begin{aligned} S_{Y_{\text{pop}}}(0) &= P(M = 0) + \sum_{m=1}^{\infty} S_{\tilde{Y}}(0)^m P(M = m) \\ &= P(M = 0) + \sum_{m=1}^{\infty} P(M = m) \\ &= \sum_{m=0}^{\infty} P(M = m) \\ &= 1, \\ \lim_{y \rightarrow \infty} S_{Y_{\text{pop}}}(y) &= P(M = 0) + \sum_{m=1}^{\infty} \lim_{y \rightarrow \infty} S_{\tilde{Y}}(y)^m P(M = m) \\ &= P(M = 0) \\ &= p_0, \end{aligned}$$

em que $p_0 = P(M = 0)$ representa a fração de cura na população.

A função de sobrevivência populacional em (2.2) pode ser reescrita como

$$\begin{aligned}
 S_{Y_{\text{pop}}}(y) &= P(M = 0) + \sum_{m=1}^{\infty} S_{\hat{Y}}(y)^m P(M = m) \\
 &= \sum_{m=0}^{\infty} S_{\hat{Y}}(y)^m P(M = m) \\
 &= E[S_{\hat{Y}}(y)^M] \\
 &= M_M [\log\{S_{\hat{Y}}(y)\}], \tag{2.3}
 \end{aligned}$$

em que $M_M(t)$ denota a função geradora de momentos da variável aleatória M , avaliada em t . Assim, para diferentes distribuições de M com função geradora de momentos existente em $\log(S_{\hat{Y}}(y))$, obtemos diferentes modelos. Por exemplo, os modelos de mistura e de tempo de promoção, quando $M \sim \text{Ber}(\theta)$ e $M \sim \text{Poisson}(\theta)$, respectivamente.

(i) Se $M \sim \text{Ber}(\theta)$, com $\theta \in (0, 1)$, a FSP em (2.3) reduz-se ao modelo de mistura com

$$\begin{aligned}
 S_{Y_{\text{pop}}}(y) &= M_M [\log\{S_{\hat{Y}}(y)\}] \\
 &= (1 - \theta) + \theta \exp\{\log[S_{\hat{Y}}(y)]\} \\
 &= (1 - \theta) + \theta S_{\hat{Y}}(y).
 \end{aligned}$$

(ii) Se $M \sim \text{Poisson}(\theta)$, com $\theta \in (0, \infty)$, a FSP em (2.3) reduz-se ao modelo de tempo de promoção com

$$\begin{aligned}
 S_{Y_{\text{pop}}}(y) &= M_M [\log\{S_{\hat{Y}}(y)\}] \\
 &= \exp\{\theta (\exp\{\log[S_{\hat{Y}}(y)]\} - 1)\} \\
 &= \exp\{-\theta F_{\hat{Y}}(y)\}.
 \end{aligned}$$

(iii) Se $M \sim \text{Geom}(\theta)$, com $\theta \in (0, 1)$, a FSP em (2.3) reduz-se ao modelo geométrico com

$$\begin{aligned}
 S_{Y_{\text{pop}}}(y) &= M_M [\log\{S_{\hat{Y}}(y)\}] \\
 &= \frac{\theta}{1 - (1 - \theta) \exp\{\log[S_{\hat{Y}}(y)]\}} \\
 &= \frac{\theta}{1 - (1 - \theta) S_{\hat{Y}}(y)},
 \end{aligned}$$

em que $p_0 = \lim_{y \rightarrow \infty} \frac{\theta}{1 - (1 - \theta) S_{\hat{Y}}(y)} = \theta$ representa a fração de cura no modelo geométrico.

A condição temporal do modelo em (2.1), de que o tempo é o tempo mínimo entre os indivíduos suscetíveis, pode ser flexibilizada. Cooner et al. (2007) na sua generalização do modelo tempo de promoção consideram o tempo dos indivíduos suscetíveis de forma mais flexível, em que é considerada uma variável que define quantos fatores latentes precisam atingir o limiar para que o evento ocorra.

Seja R uma variável aleatória representando o limiar do número máximo de fatores latentes necessários para a ocorrência do evento de interesse, por exemplo, o número máximo de células metastáticas ativas necessárias para a formação de um tumor, de modo que $1 \leq R \leq M$. Suponha, adicionalmente, as hipóteses do modelo anterior e que R e $\tilde{Y}_i$ são independentes. O tempo até o evento de interesse pode ser escrito como

$$Y_{pop} = \infty \times I(M = 0) + \tilde{Y}_{(R)} \times I(M \geq 1). \quad (2.4)$$

Portanto, a função de sobrevivência populacional do modelo é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= \sum_{m=0}^{\infty} \sum_{r=1}^m P(Y_{pop} > y, M = m, R = r) \\ &= \sum_{m=0}^{\infty} \sum_{r=1}^m P(Y_{pop} > y | M = m, R = r) P(R = r | M = m) P(M = m) \\ &= P(Y_{pop} > y | M = 0, R = 0) P(R = 0 | M = 0) P(M = 0) \\ &\quad + \sum_{m=1}^{\infty} \sum_{r=1}^m P(Y_{pop} > y | M = m, R = r) P(R = r | M = m) P(M = m) \\ &= P_M(0) + \sum_{m=1}^{\infty} \sum_{r=1}^m P(\tilde{Y}_{(r)} > y | M = m, R = r) P(R = r | M = m) P(M = m) \\ &= P_M(0) + \sum_{m=1}^{\infty} \sum_{r=1}^m S_{\tilde{Y}_{(r)}}(y) P_{R|M}(r) P_M(m). \end{aligned}$$

Seja K o número de variáveis $\tilde{Y}_1, \dots, \tilde{Y}_m$ que são menores ou iguais a y , para y fixo. Considerando $\tilde{Y}_r \leq y$ como sucesso, temos que $K \sim \text{Bin}(m, p)$, com $p = P(\tilde{Y}_r \leq y)$. Assim,

$$\begin{aligned} P(\tilde{Y}_{(r)} \leq y) &= P(K \geq r) = \sum_{k=r}^m \binom{m}{k} F_{\tilde{Y}_r}(y)^k S_{\tilde{Y}_r}(y)^{m-k} \\ &= 1 - \sum_{k=0}^{r-1} \binom{m}{k} F_{\tilde{Y}_r}(y)^k S_{\tilde{Y}_r}(y)^{m-k}. \end{aligned}$$

Logo,

$$\begin{aligned} P(\tilde{Y}_{(r)} > y) &= \sum_{k=0}^{r-1} \binom{m}{k} F_{\tilde{Y}_r}(y)^k S_{\tilde{Y}_r}(y)^{m-k} \\ &= \int_0^{S_{\tilde{Y}_r}(y)} m \binom{m-1}{r-1} u^{m-r} (1-u)^{r-1} du. \end{aligned}$$

Substituindo esse resultado na FSP, obtemos

$$\begin{aligned}
 S_{Y_{pop}}(y) &= P_M(0) + \sum_{m=1}^{\infty} \sum_{r=1}^m m \binom{m-1}{r-1} \int_0^{S_{\tilde{Y}_r}(y)} u^{m-r} (1-u)^{r-1} du P_{R|M}(r) P_M(m) \\
 &= P_M(0) + \sum_{m=1}^{\infty} \sum_{r=1}^m IB(S_{\tilde{Y}_r}(y); m-r+1, r) P_{R,M}(r, m) \\
 &= P_M(0) + E_{R,M}[IB(S_{\tilde{Y}_r}(y); M-R+1, R) I(M \geq R \geq 1)], \tag{2.5}
 \end{aligned}$$

em que $IB(;;)$ representa a função beta incompleta regularizada. Assim, diferentes distribuições para M e R produzem diferentes modelos.

Analisando suas propriedades, temos

$$\begin{aligned}
 S_{Y_{pop}}(0) &= P_M(0) + E_{R,M}[IB(1; M-R+1, R) \times I(M \geq R \geq 1)] \\
 &= P_M(0) + E_{R,M}[I(M \geq R \geq 1)] \\
 &= 1, \\
 \lim_{t \rightarrow \infty} S_{Y_{pop}}(t) &= P_M(0) + E_{R,M}[IB(0; M-R+1, R) \times I(M \geq R \geq 1)] \\
 &= P_M(0) \\
 &= p_0,
 \end{aligned}$$

em que $p_0 = P(M=0)$ representa a fração de cura na população.

Esse modelo antecede o modelo em (2.1), de fato, o engloba como caso particular quando R segue uma distribuição degenerada em 1:

Seja $R \sim \text{Degenerada}(1)$. Obtém-se o modelo com esquema de primeira ativação. Isso significa que apenas um fator latente é necessário para provocar o evento de interesse

$$\begin{aligned}
 S_{Y_{pop}}(y) &= P_M(0) + E_M[IB(S_{\tilde{Y}_r}(y); M, 1) I(M \geq 1)] \\
 &= P_M(0) + E_M[S_{\tilde{Y}_r}(y)^M I(M \geq 1)] \\
 &= P_M(0) + \sum_{m=1}^{\infty} S_{\tilde{Y}_r}(y)^m P_M(m) \\
 &= E_M[S_{\tilde{Y}_r}(y)^M].
 \end{aligned}$$

Seja $R \sim \text{Degenerada}(M)$. Obtém-se o modelo com esquema de última ativação, ou seja, o evento só ocorre quando todos os fatores latentes tiverem atingido o limiar

$$\begin{aligned}
 S_{Y_{pop}}(y) &= P_M(0) + E_M[IB(S_{\tilde{Y}_r}(y); 1, M) \times I(M \geq 1)] \\
 &= P_M(0) + E_M[1 - F_{\tilde{Y}_r}(y)^M \times I(M \geq 1)] \\
 &= P_M(0) + 1 - E_M[F_{\tilde{Y}_r}(y)^M],
 \end{aligned}$$

em que $p_0 = \lim_{y \rightarrow \infty} P_M(0) + 1 - E_M[F_{\tilde{Y}_r}(y)^M] = P_M(0)$ representa a fração de cura no esquema de última ativação.

2.4 ESQUEMA DE ATIVAÇÃO ALEATÓRIA

O modelo em (2.4) permite que o número de fatores necessários seja estocástico, dependendo do valor de M , como no caso em que R segue uma distribuição Uniforme discreta de 1 a M , denotada por $R | M \sim U\{1 \dots M\}$. Considerando $r^* = r - 1$, obtém-se o modelo com esquema de ativação aleatória, pois o número de fatores latentes necessários para o evento é aleatório e pode assumir qualquer valor entre 1 e M , com igual probabilidade. A função de sobrevivência populacional associada a esse modelo é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= P_M(0) + \sum_{m=1}^{\infty} \int_0^{S_{\tilde{Y}_r}(y)} \sum_{r^*=0}^{m-1} \binom{m-1}{r^*} u^{m-r^*-1} (1-u)^{r^*} du P_M(m) \\ &= P_M(0) + \sum_{m=1}^{\infty} \int_0^{S_{\tilde{Y}_r}(y)} du P_M(m) \\ &= P_M(0) + \sum_{m=1}^{\infty} S_{\tilde{Y}_r}(y) P_M(m) \\ &= P_M(0) + S_{\tilde{Y}_r}(y) P(M \geq 1) \\ &= P_M(0) + [1 - P_M(0)] S_{\tilde{Y}_r}(y), \end{aligned}$$

em que $p_0 = \lim_{y \rightarrow \infty} S_{Y_{pop}}(y) = P_M(0)$ representa a fração de cura no esquema de ativação aleatória.

Nota-se que esse modelo coincide com os modelos Bernoulli e geométrico de primeira ativação, quando $M \sim \text{Ber}(\theta)$ e $M \sim \text{Geom}(1 - \theta)$. Nesses casos, os esquemas de primeira ativação e ativação aleatória tornam-se equivalentes.

A partir do esquema de ativação aleatória, é possível obter diferentes modelos dependendo da distribuição de M . Por exemplo

(i) Se $M \sim \text{Ber}(\theta)$, a FSP é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= P_M(0) + [1 - P_M(0)] S_{\tilde{Y}_r}(y) \\ &= 1 - \theta + \theta S_{\tilde{Y}_r}(y), \end{aligned}$$

com fração de cura $p_0 = 1 - \theta$.

(ii) Se $M \sim \text{Poisson}(\theta)$, a **FSP** é dada por

$$S_{Y_{pop}}(y) = \exp\{-\theta\} + (1 - \exp\{-\theta\}) S_{\tilde{Y}_r}(y),$$

com fração de cura $p_0 = \exp\{-\theta\}$.

(iii) Se $M \sim \text{Geom}(\theta)$, a **FSP** é dada por

$$S_{Y_{pop}}(y) = \theta + (1 - \theta) S_{\tilde{Y}_r}(y),$$

com fração de cura $p_0 = \theta$.

Kim, Chen e Dey (2011) flexibilizam a suposição sobre o suporte da variável que representa o número de fatores necessários para a ocorrência do evento de interesse, permitindo que esse número seja maior que o total de fatores presentes. Nesse contexto, o número de fatores necessários pode ser interpretado como o nível de defesa contra a ocorrência do evento.

Ao retirar a suposição de dependência entre R e M , um novo modelo é proposto, em que R representa, por exemplo, o nível de anticorpos do sistema imunológico, assumindo que $R \geq 1$. As demais suposições permanecem as mesmas, e o tempo até o evento de interesse pode ser escrito como

$$Y_{pop} = \infty \times I(M < R) + \tilde{Y}_{(R)} \times I(M \geq R). \quad (2.6)$$

Assim, a função de sobrevivência populacional do modelo é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} P(Y_{pop} > y, M = m, R = r) \\ &= \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} P(Y_{pop} > y \mid M = m, R = r) P(M = m, R = r) \\ &= \sum_{m=0}^{\infty} \sum_{r=m+1}^{\infty} P(Y_{pop} > y \mid M = m, R = r) P(R = r, M = m) \\ &\quad + \sum_{m=0}^{\infty} \sum_{r=1}^m P(Y_{pop} > y \mid M = m, R = r) P(R = r, M = m) \\ &= \sum_{m=0}^{\infty} \sum_{r=m+1}^{\infty} P_{R,M}(r, m) + P(Y_{pop} > y \mid M = 0, R = 0) P(R = 0, M = 0) \\ &\quad + \sum_{m=1}^{\infty} \sum_{r=1}^m P(Y_{pop} > y \mid M = m, R = r) P(R = r, M = m) \\ &= P(R \geq M + 1) + P(M = 0 \mid R = 0) P(R = 0) \\ &\quad + \sum_{m=1}^{\infty} \sum_{r=1}^m S_{\tilde{Y}_{(r)}} P(R = r, M = m) \end{aligned}$$

$$\begin{aligned}
&= P(M < R) + \sum_{m=1}^{\infty} \sum_{r=1}^m IB(S_{\tilde{Y}_r}; m - r + 1, r) P(R = r, M = m) \\
&= P(M < R) + E_{R,M} \left[IB(S_{\tilde{Y}_r}; M - R + 1, R) I(M \geq R \geq 1) \right].
\end{aligned}$$

Analisando suas propriedades, temos

$$\begin{aligned}
S_{Y_{pop}}(0) &= P(M < R) + E_{R,M} \left[IB(S_{\tilde{Y}_r}(0); M - R + 1, R) I(M \geq R \geq 1) \right] \\
&= P(M < R) + E_{R,M} \left[IB(1; M - R + 1, R) I(M \geq R \geq 1) \right] \\
&= P(M < R) + E_{R,M} \left[I(M \geq R \geq 1) \right] \\
&= 1,
\end{aligned}$$

$$\begin{aligned}
\lim_{t \rightarrow \infty} S_{Y_{pop}}(t) &= P(M < R) + E_{R,M} \left[IB \left(\lim_{t \rightarrow \infty} S_{\tilde{Y}_r}(t); M - R + 1, R \right) \times I(M \geq R \geq 1) \right] \\
&= P(M < R) + E_{R,M} \left[IB(0; M - R + 1, R) \times I(M \geq R \geq 1) \right] \\
&= P(M < R) + E_{R,M} \left[0 \times I(M \geq R \geq 1) \right] \\
&= P(M < R) \\
&= p_0,
\end{aligned}$$

em que $p_0 = P(M < R)$ representa a fração de cura na população. Nota-se que, embora a contribuição dos indivíduos não curados seja igual à do modelo de [Cooner et al. \(2007\)](#), a fração de cura difere, permitindo que a cura não esteja restrita à ausência de fatores latentes que competem para a ocorrência do evento. Ela pode ocorrer mesmo quando tais fatores estão presentes, mas em quantidade insuficiente para causar o evento.

Como o suporte de R não está mais limitado pelo valor de M , não há mais o conceito de última ativação, diferentemente da primeira ativação, que se mantém.

Seja $R \sim \text{Degenerada}(1)$. Obtém-se o modelo com esquema de primeira ativação, cuja **FSP** é dada por

$$\begin{aligned}
S_{Y_{pop}}(y) &= P(M < 1) + E_M \left[IB(S_{\tilde{Y}_r}(y); M, 1) I(M \geq 1) \right] \\
&= P_M(0) + E_M \left[S_{\tilde{Y}_r}(y)^M I(M \geq 1) \right] \\
&= P_M(0) + \sum_{m=1}^{\infty} S_{\tilde{Y}_r}(y)^m P_M(m) \\
&= E_M \left[S_{\tilde{Y}_r}(y)^M \right].
\end{aligned}$$

2.5 ESQUEMA DE ATIVAÇÃO GEOMÉTRICA

O modelo em (2.6) permite que o número de fatores necessários seja estocástico, sem depender da distribuição do número total de fatores. Por exemplo, quando $R \sim \text{Geom}(\xi)$ e assumindo independência entre R e M , chega-se ao modelo com esquema de ativação geométrica, pois o número de fatores latentes segue uma distribuição geométrica. A função de sobrevivência populacional é dada por

$$\begin{aligned}
 S_{pop}(y) &= P(M < R) + E_{M,R} \left[IB \left(S_{\tilde{Y}_r}(y); M - R + 1, R \right) I(M \geq R) \right] \\
 &= \sum_{m=0}^{\infty} \sum_{r=m+1}^{\infty} P_R(r) P_M(m) \\
 &\quad + \sum_{m=1}^{\infty} m \int_0^{S_{\tilde{Y}_r}(y)} \sum_{r=1}^m \binom{m-1}{r-1} u^{m-r} (1-u)^{r-1} P_{R,M}(r, m) du \\
 &= \sum_{m=0}^{\infty} \sum_{r=m+1}^{\infty} (1-\xi)^{r-1} \xi P_M(m) \\
 &\quad + \sum_{m=1}^{\infty} m \int_0^{S_{\tilde{Y}_r}(y)} \sum_{k=0}^{m-1} \binom{m-1}{k} u^{m-1-k} (1-u)^k P_R(k+1) P_M(m) du \\
 &= \sum_{m=0}^{\infty} (1-\xi)^m P_M(m) \\
 &\quad + \xi \sum_{m=1}^{\infty} m \int_0^{S_{\tilde{Y}_r}(y)} \sum_{k=0}^{m-1} \binom{m-1}{k} u^{m-1-k} [(1-u)(1-\xi)]^k P_M(m) du \\
 &= M_M[\log(1-\xi)] + \xi \sum_{m=1}^{\infty} m \int_0^{S_{\tilde{Y}_r}(y)} [1-\xi(1-u)]^{m-1} P_M(m) du.
 \end{aligned}$$

Fazendo $x = 1 - \xi(1-u)$ e $dx = \frac{du}{\xi}$, temos

$$\begin{aligned}
 S_{pop}(y) &= M_M[\log(1-\xi)] + \sum_{m=1}^{\infty} m \int_{1-\xi}^{1-\xi F_{\tilde{Y}_r}(y)} x^{m-1} P_M(m) dx \\
 &= M_M[\log(1-\xi)] + \sum_{m=1}^{\infty} \left[(1-\xi F_{\tilde{Y}_r}\{y\})^m - (1-\xi)^m \right] P_M(m) \\
 &= M_M[\log(1-\xi)] + E_M[(1-\xi F_{\tilde{Y}_r}\{y\})^M] - E_M[(1-\xi)^M] \\
 &= M_M[\log(1-\xi F_{\tilde{Y}_r}\{y\})].
 \end{aligned}$$

Nota-se que este modelo é igual ao modelo de Rodrigues et al. (2009) quando $\xi = 1$, ou seja, quando $R \sim \text{Geom}(1)$. Nesse caso, os esquemas de primeira ativação e ativação geométrica coincidem, o que é esperado, pois a probabilidade do evento ocorrer na primeira ativação seria 1.

Do esquema de ativação geométrica, podemos derivar diferentes modelos, dependendo da distribuição de M . Por exemplo,

(i) Se $M \sim \text{Ber}(\theta)$, a **FSP** é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= M_M[\log\{1 - \xi F_{\tilde{Y}_r}(y)\}] \\ &= (1 - \theta) + \theta \exp\{\log\{1 - \xi F_{\tilde{Y}_r}(y)\}\} \\ &= (1 - \theta) + \theta[1 - \xi F_{\tilde{Y}_r}(y)], \end{aligned}$$

em que a fração de cura é $p_0 = 1 - \theta\xi$.

(ii) Se $M \sim \text{Poisson}(\theta)$, a **FSP** é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= M_M[\log\{1 - \xi F_{\tilde{Y}_r}(y)\}] \\ &= \exp\left\{\theta \left(\exp\{\log[1 - \xi F_{\tilde{Y}_r}(y)]\} - 1\right)\right\} \\ &= \exp\left\{\theta(1 - \xi F_{\tilde{Y}_r}(y) - 1)\right\} \\ &= \exp\left\{-\theta\xi F_{\tilde{Y}_r}(y)\right\}, \end{aligned}$$

em que a fração de cura é $p_0 = \exp\{-\theta\xi\}$.

(iii) Se $M \sim \text{Geom}(\theta)$, a **FSP** é dada por

$$\begin{aligned} S_{Y_{pop}}(y) &= M_M[\log\{1 - \xi F_{\tilde{Y}_r}(y)\}] \\ &= \frac{\theta}{1 - (1 - \theta) \exp\{\log[1 - \xi F_{\tilde{Y}_r}(y)]\}} \\ &= \frac{\theta}{1 - [1 - \theta][1 - \xi F_{\tilde{Y}_r}(y)]}, \end{aligned}$$

em que a fração de cura é $p_0 = \frac{\theta}{\theta + (1 - \theta)\xi}$.

3 MODELO DE FRAÇÃO DE CURA LOG-SIMÉTRICO COM ESQUEMAS DE ATIVAÇÃO

Antes de definir a classe de distribuições **LS**, é necessário apresentar a classe da qual ela se origina, a classe de distribuições simétricas, amplamente estudada na literatura (**FANG, 1990**).

Definição 3. Seja Y uma variável aleatória contínua. Dizemos que Y pertence a classe de distribuições simétricas, e denotamos $Y \sim S(\mu, \phi, g(\cdot))$, com parâmetros de localização $\mu \in \mathbb{R}$, escala $\phi \in (0, \infty)$ e função geradora de densidade $g(\cdot)$, quando a sua **FDP** é dada por

$$f_Y(y) = \frac{g\left[\frac{(y-\mu)^2}{\phi}\right]}{\sqrt{\phi}}, y \in \mathbb{R}.$$

3.1 DISTRIBUIÇÃO LOG-SIMÉTRICA

A classe de distribuições log-simétricas é definida por meio da transformação $T = \exp\{Y\}$, em que Y pertence à classe de distribuições simétricas, de forma que

$$f_T(t) = \frac{g\left(\frac{\tilde{t}^2}{t\sqrt{\phi}}\right)}{t\sqrt{\phi}}, \quad (3.1)$$

em que $\tilde{t} = \log\left[\left(\frac{t}{\eta}\right)^{\frac{1}{\sqrt{\phi}}}\right]$ e $\eta = \exp\{\mu\}$.

Com isso, tem-se uma nova variável aleatória contínua, assimétrica e com suporte no intervalo $(0, \infty)$, denotada por $T \sim LS(\eta, \phi, g(\cdot))$. Essa classe é caracterizada pelos parâmetros de localização $\eta > 0$, de potência $\phi > 0$ e função geradora de densidade $g(\cdot)$, a qual pode depender de um ou mais parâmetros adicionais, denotados por ζ , assumidos como conhecidos neste trabalho.

A de T pode ser expressa em termos da de uma distribuição simétrica, conforme

$$F_T(t) = F_Z(\tilde{t}),$$

em que $F_Z(\cdot)$ é a da variável padronizada $Z = (Y - \mu)/\sqrt{\phi} \sim S(0, 1, g(\cdot))$.

A depender da escolha da função geradora de densidade $g(\cdot)$, muitas distribuições conhecidas podem ser expressas como casos particulares da classe de distribuições **LS**. Algumas delas são apresentadas na Tabela 1.

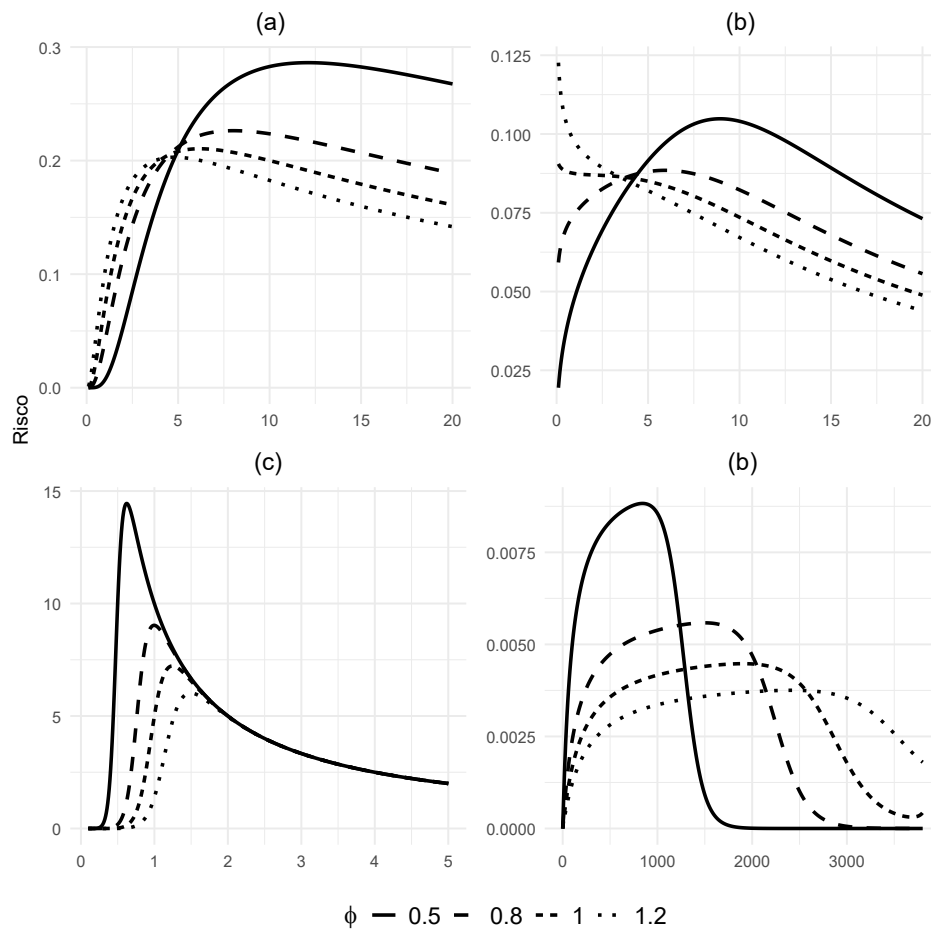
Tabela 1 – Função geradora para alguns casos particulares da classe log-simétrica.

Distribuição	Função geradora de densidade	Intervalo de ζ
Log-normal	$\exp\left\{-\frac{1}{2}u\right\}$	-
Log-t-Student	$\left(1 + \frac{u}{\zeta}\right)^{-\frac{\zeta+1}{2}}$	$\zeta > 0$
Log-hiperbólica	$\exp\left(-\zeta\sqrt{1+u}\right)$	$\zeta > 0$
Log-slash	$FGII^1\left(\zeta + \frac{1}{2}, \frac{u}{2}\right)$	$\zeta > 0$
Log-normal contaminada	$\sqrt{\zeta_2} \exp\left(-\frac{1}{2}\zeta_2 u\right) + \frac{(1-\zeta_1) \exp\left(-\frac{1}{2}u\right)}{\zeta_1}$	$0 < \zeta_1, \zeta_2 < 1$
Birnbaum-Saunders extendida	$\exp\left(-\frac{2}{\zeta^2} \sinh^2(\sqrt{u})\right) \frac{1}{\zeta \cosh(\sqrt{u})}$	$\zeta > 0$
Log-logística I	$c \frac{e^{-u}}{(1+e^{-u})^2}, c \approx 1.4843$	-
Log-logística II	$\frac{e^{-\sqrt{u}}}{(1+e^{-\sqrt{u}})^2}$	-
Log-exp. potência	$\frac{1}{C(\zeta)} \exp\left(-\frac{1}{2}u^{\frac{1}{1+\zeta}}\right)$	$-1 < \zeta \leq 1$

¹ Função gama incompleta inferior.

Fonte: o autor.

Figura 1 – Gráficos das funções de risco para distribuições log-normal(η, ϕ)(a), log-hiperbólica($\eta, \phi, \zeta=1$)(b), log-logística II(η, ϕ)(c) e Harmonic Law estendida($\eta, \phi, \zeta = 2$)(d) para diferentes valores de ϕ e $\eta = 10$.



Fonte: o autor

A flexibilidade da classe de distribuições **LS** é refletida no comportamento de sua função de risco, a qual pode assumir diferentes formatos, como mostrado na Figura 1 em que, por exemplo, as curvas apresentam comportamento decrescente, crescente, assimétrico e banheira inversa. Além de apresentar grande flexibilidade ao incluir diversas distribuições conhecidas como casos particulares, a classe **LS** possui propriedades importantes, no que segue destacamos duas dessas propriedades, as quais serão úteis para o desenvolvimento deste trabalho.

Propriedade 1. O quantil de ordem p da variável $T \sim LS(\eta, \phi, g(\cdot))$ é dado por

$$\tau_p = \eta \exp \left\{ \sqrt{\phi} Z_{\zeta}^{(p)} \right\}, \quad (3.2)$$

em que $Z_{\zeta}^{(p)}$ representa o quantil de ordem p da distribuição padronizada $Z = (Y - \mu)/\sqrt{\phi} \sim S(0, 1, g(\cdot))$.

Propriedade 2. Se $T \sim LS(\eta, \phi, g(\cdot))$, então a mediana de T é dada por

$$\text{Mediana}(T) = \eta.$$

Propriedades adicionais podem ser encontradas em Vanegas e Paula (2016).

3.2 MODELOS DE FRAÇÃO DE CURA LOG-SIMÉTRICOS

(ROCHA; MEDEIROS; VALENÇA, 2022) propuseram um modelo de fração de cura log-simétrico. Nesse modelo, o mecanismo do tempo até o evento de interesse corresponde ao mesmo mecanismo do esquema de primeira ativação, ou seja, o tempo até o evento é o tempo até o primeiro fator latente atingir um limiar, sendo que o tempo até atingir esse limiar, para cada fator latente, segue a classe de distribuições **LS**. Assim, sua **FSP** é dada por

$$S_{Y_{\text{pop}}}(y) = P(M = 0) + \sum_{m=1}^{\infty} S_{\tilde{Y}}(y)^m P(M = m)$$

onde $S_{\tilde{Y}}(y) = 1 - F_Z(\tilde{y})$, com $\tilde{Y} \sim LS(\eta, \phi, g(\cdot))$.

Esse modelo pode ser generalizado, flexibilizando o mecanismo do tempo até o evento de interesse do esquema de primeira ativação, de modo a contemplar uma quantidade maior de casos práticos, permitindo modelar diferentes mecanismos naturais dos dados. Dessa forma, é proposto o modelo de fração de cura log-simétrico com diferentes esquemas de ativação, cuja **FSP** é dada por

$$S_{Y_{\text{pop}}}(y) = P(M < R) + E_{R,M} \left[IB(S_{\tilde{Y}_r}(y); M - R + 1, R) I(M \geq R \geq 1) \right],$$

onde $S_{\tilde{Y}}(y) = 1 - F_Z(\tilde{y})$, com $\tilde{Y} \sim LS(\eta, \phi, g(\cdot))$.

3.3 INFERÊNCIA

Seja uma amostra de tamanho n , em que o tempo até a ocorrência do evento de interesse T_i é gerado por M_i fatores que competem para ocasionar esse evento, cada um associado a um tempo $\tilde{Y}_i$. Considera-se ainda um nível de defesa R_i contra o evento de interesse, também chamada de distribuição do esquema de ativação, com parâmetro ξ associado, de modo que $T_i = \min(Y_i, C_i)$, em que $Y_i = \tilde{Y}_{i(R_i)}$, para $R_i \geq 1$, e C_i representa o tempo de censura do indivíduo. Define-se o indicador da ocorrência do evento de interesse δ_i como

$$\delta_i = \begin{cases} 1, & \text{se } Y_i \leq C_i, \\ 0, & \text{se } Y_i > C_i, \end{cases} \quad \text{para } i = 1, \dots, n.$$

Considera-se $\tilde{Y}_{i(R_i)} \stackrel{\text{i.i.d.}}{\sim} LS(\eta, \phi, g(\cdot))$ sendo referida como distribuição da latência e que existe um vetor de covariáveis $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^\top$ associado ao i -ésimo indivíduo, com $x_{i1} = 1$.

Neste estudo, o parâmetro da distribuição do nível de defesa contra o evento de interesse é considerado constante em toda a amostra e denotado por ξ . Um exemplo seria um estudo clínico sobre o tempo até a recidiva de um tipo de câncer, em que os pacientes têm as mesmas características clínicas associadas ao nível do sistema imunológico.

Para incorporar o efeito das covariáveis na função de sobrevivência populacional e investigar possíveis efeitos na cura dos indivíduos, considera-se que o parâmetro θ_i , associado à distribuição do número de fatores competidores M_i , também referida como distribuição da incidência, pode variar entre os indivíduos, como exemplo, um câncer pode ser causado por um número de células tumorais que resistiram ao tratamento e estão crescendo até se tornar um tumor detectável, com os pacientes que passaram por diferentes tratamentos apresentando uma quantidade diferente de células tumorais. Dessa forma, θ_i poderá variar na amostra por meio de uma função de ligação b , de modo que

$$\theta_i = b(\mathbf{x}_i^\top \boldsymbol{\beta}),$$

sendo $\beta = (\beta_0, \beta_1, \dots, \beta_p)^\top$ o vetor de parâmetros de regressão a ser estimado.

A escolha da função de ligação deve ser compatível com o suporte do parâmetro da distribuição. Por exemplo, para a distribuição de Poisson, pode-se utilizar a ligação logarítmica, enquanto para as distribuições Bernoulli e geométrica, a ligação logística pode ser adequada. Essas ligações são dadas, respectivamente, por

$$\log(\theta_i) = \mathbf{x}_i^\top \beta$$

e

$$\log\left(\frac{\theta_i}{1 - \theta_i}\right) = \mathbf{x}_i^\top \beta.$$

Assim, obtém-se, respectivamente:

$$\theta_i = \frac{\exp\{\mathbf{x}_i^\top \beta\}}{1 + \exp\{\mathbf{x}_i^\top \beta\}} \quad (3.3)$$

e

$$\theta_i = \exp\{\mathbf{x}_i^\top \beta\}. \quad (3.4)$$

Com isso, o vetor de parâmetros a ser estimado é dado por $\lambda = (\eta, \phi, \beta, \xi)^\top$. O conjunto de dados observáveis é denotado por $D = (n, \mathbf{T}, \delta, \mathbf{X})$, enquanto o conjunto de dados completos é representado por $D_c = (n, \mathbf{T}, \delta, \mathbf{R}, \mathbf{M}, \mathbf{X})$, em que

$$\mathbf{T} = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{pmatrix}, \quad \delta = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}.$$

Utilizando o conjunto de dados completos, assumindo que o tipo de censura é aleatório e que o mecanismo de censura é não informativa, pode-se obter a função de verossimilhança completa, sabendo que

$$\begin{aligned} P(T_i = t, \delta_i = 0 | M_i, R_i) &= \frac{P(T_i = t, \delta_i = 0, M_i, R_i)}{P(M_i, R_i)} \\ &= \frac{P(C_i = t, Y_i > C_i, M_i, R_i)}{P(M_i, R_i)} \\ &= \frac{P(Y_i > t | M_i, R_i) P(C_i = t | M_i, R_i) P(M_i, R_i)}{P(M_i, R_i)} \\ &= P(Y_i > t | M_i, R_i) P(C_i = t) \\ &= S_{Y_i}(t | m_i, r_i) \end{aligned}$$

e

$$\begin{aligned}
 P(T_i = t, \delta_i = 1 | M_i, R_i) &= \frac{P(T_i = t, \delta_i = 1, M_i, R_i)}{P(M_i, R_i)} \\
 &= \frac{P(Y_i = t, Y_i \leq C_i, M_i, R_i)}{P(M_i, R_i)} \\
 &= \frac{P(C_i \geq t)P(Y_i = t | M_i, R_i)P(M_i, R_i)}{P(M_i, R_i)} \\
 &= P(C_i \geq t)P(Y_i = t | M_i, R_i) \\
 &= f_{Y_i}(t | m_i, r_i).
 \end{aligned}$$

Com isso, é possível expressar a função de verossimilhança completa por meio de uma distribuição condicional da forma

$$f(t_i, \delta_i | m_i, r_i) = S_{Y_i}(t_i | m_i, r_i)^{1-\delta_i} f_{Y_i}(t_i | m_i, r_i)^{\delta_i},$$

a qual foi utilizada para construir a função de verossimilhança completa do modelo de fração de cura, da seguinte forma

$$\begin{aligned}
 L(\boldsymbol{\lambda}; \mathbf{D}_c) &= \prod_{i=1}^n f(t_i, \delta_i, m_i, r_i) \\
 &= \prod_{i=1}^n f(t_i, \delta_i | m_i, r_i) f(m_i, r_i) \\
 &= \prod_{i=1}^n S_{Y_i}(t_i | m_i, r_i)^{1-\delta_i} f_{Y_i}(t_i | m_i, r_i)^{\delta_i} f_{M_i, R_i}(m_i, r_i),
 \end{aligned}$$

em que $S_{Y_i}(t | m_i, r_i) = I(M_i \leq R_i - 1) + IB(S_{Y_i}(t_i); M_i - R_i + 1, R_i)I(M_i \geq R_i)$ e $f_{Y_i}(t_i | m_i, r_i) = f_{Y_i}(t_i | \lambda) M_i \binom{M_i - 1}{R_i - 1} S_{Y_i}(t_i | \lambda)^{M_i - R_i} F_{Y_i}(t_i | \lambda)^{R_i - 1}$.

Seu logaritmo é dado por

$$\begin{aligned}
 l(\boldsymbol{\lambda}; \mathbf{D}_c) &= \log [L(\boldsymbol{\lambda}; \mathbf{D}_c)] \\
 &= \sum_{i=1}^n (1 - \delta_i) \log \left[I(m_i \leq r_i - 1) + IB(S_{Y_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i) \right] \quad (3.5) \\
 &\quad + \delta_i \left[\log \left\{ g \left(\tilde{t}_i^2 \right) \right\} - \log(t_i) - \frac{1}{2} \log(\phi) + \log(m_i) + \log \left\{ \binom{m_i - 1}{r_i - 1} \right\} \right. \\
 &\quad \left. + (m_i - 1) \log \{ 1 - F_{Z_i}(\tilde{t}_i) \} + (r_i - 1) \log \{ F_{Z_i}(\tilde{t}_i) \} \right] + \log \{ f_{M_i, R_i}(m_i, r_i) \}. \quad (3.6)
 \end{aligned}$$

Admitindo independência entre as distribuições de R_i e M_i , os estimadores dos parâmetros correspondem aos estimadores usuais de máxima verossimilhança de suas respectivas distribuições. Para os demais parâmetros, derivamos o logaritmo da função de verossimilhança em

(3.5) com relação ao vetor de parâmetros λ , obtendo

$$\begin{aligned} \frac{\partial l}{\partial \eta} &= \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \delta_i \left[\tilde{t}_i v(\tilde{t}_i) + \frac{(m_i - 1)g(\tilde{t}_i^2)}{1 - F_{Z_i}(\tilde{t}_i)} + \frac{(r_i - 1)g(\tilde{t}_i^2)}{F_{Z_i}(\tilde{t}_i)} \right] \\ &\quad - \frac{(1 - \delta_i)g(\tilde{t}_i^2) I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1} F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i}}{I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)}, \\ \frac{\partial l}{\partial \phi} &= \frac{1}{2\phi} \sum_{i=1}^n \delta_i \left[\tilde{t}_i v(\tilde{t}_i) - 1 + \frac{(m_i - 1) \tilde{t}_i g(\tilde{t}_i^2)}{1 - F_{Z_i}(\tilde{t}_i)} + \frac{(r_i - 1) \tilde{t}_i g(\tilde{t}_i^2)}{F_{Z_i}(\tilde{t}_i)} \right] \\ &\quad - \frac{(1 - \delta_i) \tilde{t}_i g(\tilde{t}_i^2) I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1} F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i}}{I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)}. \end{aligned}$$

Suas derivadas segundas são dadas, respectivamente, por

$$\begin{aligned} \frac{\partial^2 l}{\partial \eta^2} &= \frac{1}{\eta^2 \phi} \sum_{i=1}^n (1 - \delta_i) I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1} g^2(\tilde{t}_i^2) F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} \\ &\quad \left[\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} + \frac{\sqrt{\phi} - \tilde{t}_i v(\tilde{t}_i)}{g(\tilde{t}_i^2)} \right] \frac{[IB(S(t); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\ &\quad + \frac{I(m_i \leq r_i - 1) - F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1}}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\ &\quad + \delta_i \left\{ v(\tilde{t}_i) \left[\left(1 + \sqrt{\phi} \tilde{t}_i \right) - \left\{ v(\tilde{t}_i) \tilde{t}_i \right\}^2 \right] - \frac{2\sqrt{\phi} \eta \tilde{t}_i^2 \frac{\partial^2 g(\tilde{t}_i^2)}{\partial \eta \partial \tilde{t}_i}}{g(\tilde{t}_i^2)} \right. \\ &\quad + g(\tilde{t}_i^2) \left[\left(\sqrt{\phi} - v(\tilde{t}_i) \tilde{t}_i \right) \left(\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} \right) \right. \\ &\quad \left. \left. - \frac{r_i - 1}{F_{Z_i}^2(\tilde{t}_i)} + \frac{m_i - 1}{(1 - F_{Z_i}(\tilde{t}_i))^2} \right] \right\}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \phi^2} &= \frac{1}{4\phi^2} \sum_{i=1}^n (1 - \delta_i) I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1} \tilde{t}_i^2 g^2(\tilde{t}_i^2) F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} \\
&\quad \left[\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} + \frac{1 - \tilde{t}_i v(\tilde{t}_i)}{\tilde{t}_i g(\tilde{t}_i^2)} \right] \frac{[I(m_i \leq r_i - 1) + IB(S; m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\
&\quad \frac{-F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1}}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\
&\quad + \delta_i \left\{ 2 - \frac{4 \tilde{t}_i^2 \phi \frac{\partial^2 g(\tilde{t}_i^2)}{\partial \phi \partial (\tilde{t}_i^2)}}{g(\tilde{t}_i^2)} + \tilde{t}_i g(\tilde{t}_i^2) \left[\tilde{t}_i \left(1 - \tilde{t}_i^2 v(\tilde{t}_i) \right) \left(\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} \right) \right. \right. \\
&\quad \left. \left. - g(\tilde{t}_i^2) \left(\frac{r_i - 1}{F_{Z_i}^2(\tilde{t}_i)} - \frac{m_i - 1}{(1 - F_{Z_i}(\tilde{t}_i))^2} \right) \right] - \tilde{t}_i^2 v(\tilde{t}_i) \left[4 + \tilde{t}_i^2 v(\tilde{t}_i) \right] \right\}, \\
\frac{\partial^2 l}{\partial \eta \partial \phi} &= \frac{1}{2\eta \sqrt{\phi^3}} \sum_{i=1}^n (1 - \delta_i) I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1} \tilde{t}_i^2 g^2(\tilde{t}_i^2) F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} \\
&\quad \left[\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} + \frac{1 - v(\tilde{t}_i)}{\tilde{t}_i g(\tilde{t}_i^2)} \right] \frac{[I(m_i \leq r_i - 1) + IB(S; m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\
&\quad \frac{-F_{Z_i}(\tilde{t}_i)^{r_i - 1} [1 - F_{Z_i}(\tilde{t}_i)]^{m_i - r_i} I(m_i \geq r_i) m_i \binom{m_i - 1}{r_i - 1}}{[I(m_i \leq r_i - 1) + IB(S_{Z_i}(t_i); m_i - r_i + 1, r_i) I(m_i \geq r_i)]^2} \\
&\quad + \delta_i \left\{ -\frac{-2\eta \sqrt{\phi} \tilde{t}_i^3 \frac{\partial^2 g(\tilde{t}_i^2)}{\partial \eta \partial (\tilde{t}_i^2)}}{g(\tilde{t}_i^2)} + \tilde{t}_i g(\tilde{t}_i^2) \left[\left(\frac{1}{\tilde{t}_i} - \tilde{t}_i v(\tilde{t}_i) \right) \left(\frac{r_i - 1}{F_{Z_i}(\tilde{t}_i)} - \frac{m_i - 1}{1 - F_{Z_i}(\tilde{t}_i)} \right) \right. \right. \\
&\quad \left. \left. - g(\tilde{t}_i^2) \left(\frac{r_i - 1}{F_{Z_i}^2(\tilde{t}_i)} - \frac{m_i - 1}{(1 - F_{Z_i}(\tilde{t}_i))^2} \right) \right] - \tilde{t}_i^2 v(\tilde{t}_i) \left[2 + v(\tilde{t}_i) \right] \right\},
\end{aligned}$$

em que $v(t) = -2 \frac{\partial g(t)}{g(t^2)}$.

Essa função de verossimilhança completa envolve variáveis não observáveis (dados incompletos), o que exige um cuidado especial. Em muitos modelos com dados incompletos é comum utilizar o algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977). No entanto, em diversos casos, o passo E, que envolve o cálculo do valor esperado torna-se analiticamente complexo. Existem métodos no contexto bayesiano para aproximar esse valor esperado por meio de algoritmos MCMC, no entanto, essa abordagem pode ser computacionalmente custosa e de difícil cali-

bração, devido à alta autocorrelação e ao burn-in insuficiente. Uma abordagem alternativa, no contexto clássico, consiste em utilizar a verossimilhança marginal, somando sobre todos os possíveis valores das variáveis latentes como feito nos trabalhos de [Carneiro \(2012\)](#), [Rocha, Medeiros e Valença \(2022\)](#) e [Assunção \(2023\)](#) de forma que

$$\begin{aligned} L(\boldsymbol{\lambda}; \mathbf{D}) &= \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} \prod_{i=1}^n S_{Y_i}(t_i | m_i, r_i)^{1-\delta_i} f_{Y_i}(t_i | m_i, r_i)^{\delta_i} f_{M_i, R_i}(m_i, r_i) \\ &= \prod_{i=1}^n \sum_{m_i=0}^{\infty} \sum_{r_i=1}^{\infty} S_{Y_i}(t_i | m_i, r_i)^{1-\delta_i} f_{Y_i}(t_i | m_i, r_i)^{\delta_i} f_{M_i, R_i}(m_i, r_i). \end{aligned}$$

Para $\delta = 0$ tem-se

$$\begin{aligned} L(\boldsymbol{\lambda}; \mathbf{D}) &= \prod_{i=1}^n \sum_{m_i=0}^{\infty} \sum_{r_i=1}^{\infty} S_{Y_i}(t_i | m_i, r_i) f_{M_i, R_i}(m_i, r_i) \\ &= \prod_{i=1}^n S_{Y_i}(t_i) \end{aligned}$$

e para $\delta = 1$,

$$\begin{aligned} L(\boldsymbol{\lambda}; \mathbf{D}) &= \prod_{i=1}^n \sum_{m_i=0}^{\infty} \sum_{r_i=1}^{\infty} f_{Y_i}(t_i | m_i, r_i) f_{M_i, R_i}(m_i, r_i) \\ &= \prod_{i=1}^n f_{Y_i}(t_i). \end{aligned}$$

Logo,

$$\begin{aligned} L(\boldsymbol{\lambda}; \mathbf{D}) &= \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} \prod_{i=1}^n S_{Y_i}(t_i | m_i, r_i)^{1-\delta_i} f_{Y_i}(t_i | m_i, r_i)^{\delta_i} f_{M_i, R_i}(m_i, r_i) \\ &= \prod_{i=1}^n S_{Y_i}(t_i)^{1-\delta_i} f_{Y_i}(t_i)^{\delta_i}. \end{aligned}$$

O estimador de máxima verossimilhança $\hat{\boldsymbol{\lambda}}$ pode ser obtido resolvendo o sistema de equações $U(\hat{\boldsymbol{\lambda}}) = 0$, em que $U(\boldsymbol{\lambda})$ representa o vetor escore dado por

$$U(\boldsymbol{\lambda}) = \frac{\partial \log[L(\boldsymbol{\lambda}; \mathbf{D})]}{\partial \boldsymbol{\lambda}}.$$

Sob condições de regularidade, tem-se que $\hat{\boldsymbol{\lambda}} \sim N_{p+3}(\boldsymbol{\lambda}, I^{-1}(\boldsymbol{\lambda}))$, em que

$$I^{-1}(\boldsymbol{\lambda}) = -E \left[\frac{\partial^2 \ell(\boldsymbol{\lambda}; \mathcal{D})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} \right]$$

é a matriz de informação de Fisher esperada. Contudo, devido à presença de observações censuradas, o cálculo de $I^{-1}(\boldsymbol{\lambda})$ não é viável. Nesse cenário, a inferência é baseada na matriz de informação observada, dada por

$$\ddot{\mathcal{L}}(\boldsymbol{\lambda}) = -\frac{\partial^2 \log[L(\boldsymbol{\lambda}; \mathbf{D})]}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top}.$$

A partir da matriz $\ddot{\mathcal{L}}(\boldsymbol{\lambda})$, obtêm-se os erros-padrão (EP) dos estimadores, os quais permitem a construção de intervalos de confiança para os parâmetros de interesse. Outra medida bastante utilizada, devido à sua interpretabilidade é a razão de chances, denotada por $\hat{\psi}$. No caso de uma covariável indicadora de grupo com duas categorias, tem-se

$$\hat{\psi} = \exp(\hat{\beta}).$$

Assim, um intervalo de confiança $(1 - \alpha) \times 100\%$ para a razão de chances é dado por

$$(\hat{\psi}_I, \hat{\psi}_S) = \exp \left\{ \hat{\beta} \pm z_{(1-\alpha/2)} EP(\hat{\beta}) \right\},$$

em que $z_{(1-\alpha/2)}$ representa o quantil da distribuição normal padrão correspondente.

A seguir, serão apresentadas as verossimilhanças marginais e suas respectivas derivadas para alguns dos esquemas de fração de cura possíveis neste modelo, sob diferentes distribuições para as incidências M_i :

- (i) $M_i \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_i)$: ou seja, existe no máximo uma causa competindo para a ocorrência do evento de interesse, e essa causa estará presente com probabilidade θ_i . Considera-se a parte sistemática dada em (3.3).
- (ii) $M_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_i)$: nesse caso, o número de fatores concorrendo para a ocorrência do evento de interesse é positivo, não limitado e, em média, igual a θ_i . A parte sistemática considerada é a da equação em (3.4).
- (iii) $M_i \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_i)$: representa o número de fatores até que um deles cause a ocorrência do evento de interesse. Esse número é positivo, não limitado, e cada fator tem probabilidade θ_i de causar o evento. A parte sistemática utilizada é a mesma da equação (3.3).

3.3.1 Modelo de primeira ativação log-simétrico

O esquema de primeira ativação, no qual $R_i = 1$, considera que o nível de defesa contra o evento de interesse é igual a 1. Por exemplo, pode-se interpretar esse caso como um sistema imunológico incapaz de conter qualquer quantidade de células tumorais, de modo que, a partir da presença de um único fator, o paciente poderá ir a óbito devido ao câncer.

Nesse caso, a função de verossimilhança marginal é dada por

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(M_M \left[\log \{1 - F_{Z_i}(\tilde{t}_i)\} \right] \right)^{1-\delta_i} \left(- \frac{\partial \left(M_M \left[\log \{1 - F_{Z_i}(\tilde{t}_i)\} \right] \right)}{\partial t_i} \right)^{\delta_i}.$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_i)$ temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(1 - \theta_i + \theta_i [1 - F_{Z_i}(\tilde{t}_i)] \right)^{1-\delta_i} \left(\theta_i \frac{g(\tilde{t}_i^2)}{t_i \sqrt{\phi}} \right)^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \log [L(\boldsymbol{\lambda}; \mathbf{D})] = \sum_{i=1}^n (1 - \delta_i) \log \left\{ 1 - \theta_i + \theta_i [1 - F_{Z_i}(\tilde{t}_i)] \right\} \\ + \delta_i \left(\log(\theta_i) + \log \left[g(\tilde{t}_i^2) \right] - \log(t_i \sqrt{\phi}) \right).$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v(\tilde{t}_i) - \frac{(1 - \delta_i) \theta_i g(\tilde{t}_i^2)}{1 - \theta_i + \theta_i (1 - F_{Z_i}(\tilde{t}_i))},$$

$$\frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \delta_i \left(\tilde{t}_i^2 v(\tilde{t}_i) - 1 \right) - \frac{(1 - \delta_i) \theta_i \tilde{t}_i g(\tilde{t}_i^2)}{1 - \theta_i + \theta_i (1 - F_{Z_i}(\tilde{t}_i))}$$

e

$$\frac{\partial l}{\partial \beta_k} = \sum_{i=1}^n \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left(\delta_i + (1 - \delta_i) \frac{\theta_i F_{Z_i}(\tilde{t}_i)}{1 - \theta_i F_{Z_i}(\tilde{t}_i)} \right).$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(\exp \left\{ -\theta_i F_{Z_i}(\tilde{t}_i) \right\} \right)^{(1-\delta_i)} \left(\theta_i \frac{g(\tilde{t}_i^2)}{t_i \sqrt{\phi}} \exp \left\{ -\theta_i F_{\tilde{Y}_r}(y) \right\} \right)^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \sum_{i=1}^n \delta_i \left(\log(\theta_i) + \log \left[g(\tilde{t}_i^2) \right] - \log(t_i) - \log \left[\sqrt{\phi} \right] \right) - \theta_i F_{Z_i}(\tilde{t}_i).$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v(\tilde{t}_i) + \theta_i g(\tilde{t}_i^2),$$

$$\frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \tilde{t}_i \left[\delta_i \tilde{t}_i v(\tilde{t}_i) + \theta_i g(\tilde{t}_i^2) \right] - \delta_i,$$

$$\frac{\partial l}{\partial \beta_k} = \sum_{i=1}^n x_{ik} \left(\delta_i - \theta_i F_{Z_i}(\tilde{t}_i) \right).$$

Considerando $M_i \sim \text{Geom}(\theta_i)$ temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(\frac{\theta_i}{1 - (1 - \theta_i)(1 - F_{Z_i}(\tilde{t}_i))} \right)^{1-\delta_i} \left(\frac{g(\tilde{t}_i^2) \theta_i (1 - \theta_i)}{t_i \sqrt{\phi} [1 - (1 - \theta_i)(1 - F_{Z_i}(\tilde{t}_i))]^2} \right)^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \sum_{i=1}^n \log(\theta_i) - [2 - (1 - \delta_i)] \log \left\{ 1 - [1 - \theta_i][1 - F_{Z_i}(\tilde{t}_i)] \right\} \\ + \delta_i \left\{ \log \left[g \left(\tilde{t}_i^2 \right) \right] + \log(1 - \theta_i) - \log(t_i) - \log \left(\sqrt{\phi} \right) \right\}.$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v \left(\tilde{t}_i \right) + \frac{(1 + \delta_i)(1 - \theta_i) g \left(\tilde{t}_i^2 \right)}{\theta_i + (1 - \theta_i) \left(1 - F_{Z_i}(\tilde{t}_i) \right)},$$

$$\frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \delta_i \left(\tilde{t}_i^2 v \left(\tilde{t}_i \right) - 1 \right) + \frac{(1 + \delta_i)(1 - \theta_i) \tilde{t}_i g \left(\tilde{t}_i^2 \right)}{\theta_i + (1 - \theta_i) \left(1 - F_{Z_i}(\tilde{t}_i) \right)}$$

e

$$\frac{\partial l}{\partial \beta_k} = \sum_{i=1}^n \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left(1 - (1 + \delta_i) \frac{\theta_i(1 - F_{Z_i}(\tilde{t}_i))}{1 - (1 - \theta_i)(1 - F_{Z_i}(\tilde{t}_i))} - \frac{\delta_i \theta_i}{1 - \theta_i} \right).$$

3.3.2 Modelo de ativação aleatória log-simétrico

O esquema de ativação aleatória considera que o nível de defesa contra o evento de interesse segue uma distribuição uniforme discreta, com valor máximo igual ao número de fatores competindo para a ocorrência do evento. Assim, todos os possíveis níveis de defesa têm igual probabilidade, ou seja, $R_i \stackrel{\text{ind}}{\sim} U\{1 \dots M_i\}$. Como exemplo, pode-se considerar um computador com um antivírus que possui igual probabilidade de falhar diante de qualquer um dos vírus existentes.

Nesse caso, a função de verossimilhança marginal é dada por

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(P_{M_i}(M_i = 0) + [1 - P_{M_i}(M_i = 0)] S_{Z_i}(\tilde{t}_i) \right)^{1-\delta_i} \\ \times \left([P_{M_i}(M_i = 0) - 1] \frac{\partial [S_{Z_i}(\tilde{t}_i)]}{\partial t_i} \right)^{\delta_i}.$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left[1 - \theta_i + \theta_i S_{Z_i}(\tilde{t}_i) \right]^{1-\delta_i} \left[\frac{g \left(\tilde{t}_i^2 \right)}{\theta_i \sqrt{\phi}} \right]^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \log [L(\boldsymbol{\lambda}; \mathbf{D})] = \sum_{i=1}^n (1 - \delta_i) \log \left\{ 1 - \theta_i + \theta_i [1 - F_{Z_i}(\tilde{t}_i)] \right\} \\ + \delta_i \left(\log(\theta_i) + \log \left[g(\tilde{t}_i^2) \right] - \log \left[t_i \sqrt{\phi} \right] \right).$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \frac{(1 - \delta_i) \theta_i g(\tilde{t}_i^2)}{1 - \theta_i + \theta_i S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v(\tilde{t}_i), \\ \frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \tilde{t}_i \left(\frac{(1 - \delta_i) \theta_i g(\tilde{t}_i^2)}{1 - \theta_i + \theta_i S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v(\tilde{t}_i) \right) - \delta_i \\ \text{e } \frac{\partial l}{\partial \beta_k} = \sum_{i=1}^n - \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left(\delta_i + (1 - \delta_i) \frac{\theta_i F_{Z_i}(\tilde{t}_i)}{1 - \theta_i F_{Z_i}(\tilde{t}_i)} \right).$$

Considerando $M_i \sim \text{Poisson}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(\exp\{-\theta_i\} + [1 - \exp\{-\theta_i\}] S_{Z_i}(\tilde{t}_i) \right)^{1 - \delta_i} \left([1 - \exp\{-\theta_i\}] \frac{g(\tilde{t}_i^2)}{t_i \sqrt{\phi}} \right)^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \sum_{i=1}^n (1 - \delta_i) \log \left(\exp\{-\theta_i\} + [1 - \exp\{-\theta_i\}] [1 - F_{Z_i}(\tilde{t}_i)] \right) \\ + \delta_i \left(\log(1 - \exp\{-\theta_i\}) + \log \left[g(\tilde{t}_i^2) \right] - \log \left[t_i \sqrt{\phi} \right] \right).$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \frac{(1 - \delta_i)(1 - \exp\{-\theta_i\}) g(\tilde{t}_i^2)}{\exp\{-\theta_i\} + (1 - \exp\{-\theta_i\}) S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v(\tilde{t}_i), \\ \frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \tilde{t}_i \left[\frac{(1 - \delta_i)(1 - \exp\{-\theta_i\}) g(\tilde{t}_i^2)}{\exp\{-\theta_i\} + (1 - \exp\{-\theta_i\}) S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v(\tilde{t}_i) \right] - \delta_i \\ \text{e } \frac{\partial l}{\partial \beta_k} = - \sum_{i=1}^n x_{ik} \left[\frac{\delta_i \theta_i \exp\{-\theta_i\}}{1 - \exp\{-\theta_i\}} + \frac{(1 - \delta_i) \theta_i \exp\{-\theta_i\} F_{Z_i}(\tilde{t}_i)}{\exp\{-\theta_i\} + (1 - \exp\{-\theta_i\}) S_{Z_i}(\tilde{t}_i)} \right].$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left[\theta_i + (1 - \theta_i) S_{Z_i}(\tilde{t}_i) \right]^{1 - \delta_i} \left[(1 - \theta_i) \frac{g(\tilde{t}_i^2)}{t_i \sqrt{\phi}} \right]^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}_{obs}) = \sum_{i=1}^n (1 - \delta_i) \log \left\{ \theta_i + [1 - \theta_i][1 - F_{Z_i}(\tilde{t}_i)] \right\} \\ + \delta_i \left\{ \log(1 - \theta_i) + \log \left[g \left(\tilde{t}_i^2 \right) \right] - \log \left[t_i \sqrt{\phi} \right] \right\}.$$

Derivando em relação aos seus parâmetros, temos que

$$\frac{\partial l}{\partial \eta} = \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \frac{(1 - \delta_i)(1 - \theta_i)g \left(\tilde{t}_i^2 \right)}{\theta_i + (1 - \theta_i)S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v \left(\tilde{t}_i \right), \\ \frac{\partial l}{\partial \phi} = \frac{1}{2\phi} \sum_{i=1}^n \tilde{t}_i \left[\frac{(1 - \delta_i)(1 - \theta_i)g \left(\tilde{t}_i^2 \right)}{\theta_i + (1 - \theta_i)S_{Z_i}(\tilde{t}_i)} - \delta_i \tilde{t}_i v \left(\tilde{t}_i \right) \right] - \delta_i \\ \text{e } \frac{\partial l}{\partial \beta_k} = \sum_{i=1}^n \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left[\frac{\theta_i \delta_i}{1 - \theta_i} + \frac{(1 - \delta_i)\theta_i F_{Z_i}(\tilde{t}_i)}{1 - \theta_i F_{Z_i}(\tilde{t}_i)} \right].$$

3.3.3 Modelo de ativação geométrica log-simétrico

O esquema de ativação geométrica, no qual $R_i \stackrel{\text{ind}}{\sim} \text{Geom}(\xi)$, considera que o nível de defesa contra o evento de interesse é positivo e sem limite superior. Por exemplo, pode-se interpretar esse caso como um sistema imunológico capaz de lidar com uma quantidade variável de tumores; no entanto, ao surgir um número de tumores que iguale ou exceda essa capacidade, o paciente poderá ir a óbito devido ao câncer.

Nesse caso, a função de verossimilhança marginal é dada por

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(M_M \left[\log \{1 - \xi F_{Z_i}(\tilde{t}_i)\} \right] \right)^{1 - \delta_i} \left(- \frac{\partial \left(M_M \left[\log \{1 - \xi F_{Z_i}(\tilde{t}_i)\} \right] \right)}{\partial t_i} \right)^{\delta_i}.$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left\{ 1 - \theta_i + \theta_i [1 - \xi F_{Z_i}(\tilde{t}_i)] \right\}^{1 - \delta_i} \left\{ \theta_i \xi \frac{g \left(\tilde{t}_i^2 \right)}{t_i \sqrt{\phi}} \right\}^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \log [L(\boldsymbol{\lambda}; \mathbf{D})] = \sum_{i=1}^n (1 - \delta_i) \log \left\{ 1 - \theta_i + \theta_i [1 - \xi F_{Z_i}(\tilde{t}_i)] \right\} \\ + \delta_i \left\{ \log(\theta_i) + \log(\xi) + \log \left[g \left(\tilde{t}_i^2 \right) \right] - \log \left[t_i \sqrt{\phi} \right] \right\}.$$

Derivando em relação aos seus parâmetros, temos que

$$\begin{aligned}\frac{\partial l}{\partial \eta} &= \frac{1}{\eta\sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v(\tilde{t}_i) - \frac{(1-\delta_i)\theta_i \xi g(\tilde{t}_i^2)}{1-\theta_i + \theta_i(1-\xi F_{Z_i}(\tilde{t}_i))}, \\ \frac{\partial l}{\partial \phi} &= \frac{1}{2\phi} \sum_{i=1}^n \delta_i (\tilde{t}_i^2 v(\tilde{t}_i) - 1) - \frac{(1-\delta_i)\theta_i \xi \tilde{t}_i g(\tilde{t}_i^2)}{1-\theta_i + \theta_i(1-\xi F_{Z_i}(\tilde{t}_i))}, \\ \frac{\partial l}{\partial \beta_k} &= \sum_{i=1}^n \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left(\delta_i + (1-\delta_i) \frac{\xi \theta_i F_{Z_i}(\tilde{t}_i)}{1 - \xi \theta_i F_{Z_i}(\tilde{t}_i)} \right) \\ \text{e } \frac{\partial l}{\partial \xi} &= \sum_{i=1}^n \frac{\delta_i}{\xi} + (1-\delta_i) \frac{\theta_i F_{Z_i}(\tilde{t}_i)}{1 - \xi \theta_i F_{Z_i}(\tilde{t}_i)}.\end{aligned}$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_i)$, temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left(\exp \left\{ -\theta_i \xi F_{Z_i}(\tilde{t}_i) \right\} \right)^{1-\delta_i} \left(\theta_i \xi \frac{g(\tilde{t}_i^2)}{\tilde{t}_i \sqrt{\phi}} \exp \left\{ -\theta_i \xi F_{\tilde{Y}_r}(y) \right\} \right)^{\delta_i}$$

e seu logaritmo é dado por

$$l(\boldsymbol{\lambda}; \mathbf{D}) = \sum_{i=1}^n \delta_i \left\{ \log(\theta_i) + \log(\xi) + \log \left[g(\tilde{t}_i^2) \right] - \log(t_i) - \log \left[\sqrt{\phi} \right] \right\} - \theta_i \xi F_{Z_i}(\tilde{t}_i).$$

Derivando em relação aos seus parâmetros, temos que

$$\begin{aligned}\frac{\partial l}{\partial \eta} &= \frac{1}{\eta\sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v(\tilde{t}_i) + \theta_i \xi g(\tilde{t}_i^2), \\ \frac{\partial l}{\partial \phi} &= \frac{1}{2\phi} \sum_{i=1}^n \tilde{t}_i \left[\delta_i \tilde{t}_i v(\tilde{t}_i) + \theta_i \xi g(\tilde{t}_i^2) \right] - \delta_i, \\ \frac{\partial l}{\partial \beta_k} &= \sum_{i=1}^n x_{ik} \left[\delta_i - \theta_i \xi F_{Z_i}(\tilde{t}_i) \right] \\ \text{e } \frac{\partial l}{\partial \xi} &= \sum_{i=1}^n \frac{\delta_i}{\xi} - \theta_i F_{Z_i}(\tilde{t}_i).\end{aligned}\tag{3.7}$$

Considerando $M_i \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_i)$ temos que

$$L(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{i=1}^n \left\{ \frac{\theta_i}{1 - [1 - \theta_i][1 - \xi F_{Z_i}(\tilde{t}_i)]} \right\}^{1-\delta_i} \left\{ \frac{g(\tilde{t}_i^2) \theta_i (1 - \theta_i) \xi}{\tilde{t}_i \sqrt{\phi} [1 - (1 - \theta_i)(1 - \xi F_{Z_i}(\tilde{t}_i))]^2} \right\}^{\delta_i}$$

e seu logaritmo é dado por

$$\begin{aligned}l(\boldsymbol{\lambda}; \mathbf{D}) &= \sum_{i=1}^n \log(\theta_i) - [2 - (1 - \delta_i)] \log \left\{ 1 - [1 - \theta_i][1 - \xi F_{Z_i}(\tilde{t}_i)] \right\} \\ &\quad + \delta_i \left\{ \log \left[g(\tilde{t}_i^2) \right] + \log(\xi) + \log(1 - \theta_i) - \log(t_i) - \log \left(\sqrt{\phi} \right) \right\}.\end{aligned}$$

Derivando em relação aos seus parâmetros, temos que

$$\begin{aligned}
 \frac{\partial l}{\partial \eta} &= \frac{1}{\eta \sqrt{\phi}} \sum_{i=1}^n \delta_i \tilde{t}_i v\left(\tilde{t}_i\right) + \frac{(1 + \delta_i)(1 - \theta_i) \xi g\left(\tilde{t}_i^2\right)}{\theta_i + (1 - \theta_i) \left(1 - \xi F_{Z_i}(\tilde{t}_i)\right)}, \\
 \frac{\partial l}{\partial \phi} &= \frac{1}{2\phi} \sum_{i=1}^n \delta_i \left(\tilde{t}_i^2 v\left(\tilde{t}_i\right) - 1\right) + \frac{(1 + \delta_i)(1 - \theta_i) \xi \tilde{t}_i g\left(\tilde{t}_i^2\right)}{\theta_i + (1 - \theta_i) \left(1 - \xi F_{Z_i}(\tilde{t}_i)\right)}, \\
 \frac{\partial l}{\partial \beta_k} &= \sum_{i=1}^n \frac{x_{ik}}{1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \left[1 - (1 + \delta_i) \frac{\theta_i(1 - \xi F_{Z_i}(\tilde{t}_i))}{1 - (1 - \theta_i)(1 - \xi F_{Z_i}(\tilde{t}_i))} - \frac{\delta_i \theta_i}{1 - \theta_i} \right] \\
 \text{e } \frac{\partial l}{\partial \xi} &= \sum_{i=1}^n \frac{\delta_i}{\xi} + \frac{(1 + \delta_i)(1 - \theta_i) F_{Z_i}(\tilde{t}_i)}{1 - (1 - \theta_i)(1 - \xi F_{Z_i}(\tilde{t}_i))}.
 \end{aligned}$$

4 ESTUDO DE SIMULAÇÃO

4.1 ESTIMAÇÃO

Com o intuito de avaliar os estimadores dos parâmetros no modelo de fração de cura log-simétrico com diferentes esquemas de ativação, foi realizado um estudo de simulação de Monte Carlo considerando os esquemas de primeira ativação, ativação aleatória e ativação geométrica. A distribuição de incidência adotada foi a Poisson, enquanto a distribuição de latência segue as distribuições log-normal e log-t-Student com três graus de liberdade. O objetivo deste estudo foi analisar o impacto de diferentes frações de curados, tamanhos amostrais e níveis de censura para os indivíduos não curados no comportamento dos estimadores dos parâmetros da distribuição para diferentes esquemas de ativação e de distribuição de latência.

O primeiro Cenário considera uma variável explicativa x_i dicotômica representando 1 se o caso for tratamento e 0 se for controle, seguindo uma distribuição de Bernoulli com probabilidade 0,5. O segundo Cenário considera uma covariável x_i contínua, seguindo uma distribuição normal padrão e em ambos os Cenários os valores foram fixados nas diferentes réplicas da simulação. Os valores verdadeiros dos parâmetros utilizados nesta simulação foram fixados em $\eta = 5$, $\phi = 1$ e $\xi = 0,5$. Como a **fração de cura (fc)** depende do parâmetro θ , e este, por sua vez, é dado por $\theta_i = \exp\{\mathbf{x}_i^\top \beta\}$, com $\mathbf{x}_i^\top = (1, x_i)$, o vetor de coeficientes β foi ajustado de modo que a **fc**, em média, assumisse os valores desejados de 10%, 20% e 30%. Como a probabilidade de cura nos esquemas de primeira ativação e ativação aleatória é a mesma $P(M = 0)$, os vetores β utilizados nesses casos foram iguais. Já no esquema de ativação geométrica, a probabilidade de cura é dada por $P(M < R)$, diferente dos esquemas com $P(M = 0)$, então são exigidos vetores β diferentes. Esses valores estão descritos nas Tabelas 2 e 3.

Além da censura oriunda da **fc**, foi adicionada censura nos indivíduos não curados, de modo que a **proporção de censura nos não curados (pcNC)** fosse de 15% e 45%. No geral, a censura total nas simulações varia desde um nível relativamente baixo, em torno de 25%, até um nível considerado alto, de 75%. Também foram considerados diferentes tamanhos amostrais, sendo $n = 50, 150, 250$ e 500 .

Tabela 2 – Vetores fixados para β nas simulações dos esquemas de primeira ativação e ativação aleatória

fc(%)	Cenário 1	Cenário 2
10	$\beta = (0,77, 0,15)^\top$	$\beta = (1,51, 1,15)^\top$
20	$\beta = (0,40, 0,15)^\top$	$\beta = (0,86, 1,15)^\top$
30	$\beta = (0,11, 0,15)^\top$	$\beta = (0,39, 1,15)^\top$

Tabela 3 – Vetores fixados para β nas simulações do esquema de ativação geométrica

fc(%)	Cenário 1	Cenário 2
10	$\beta = (1,21, 2,00)^\top$	$\beta = (3,10, 2,00)^\top$
20	$\beta = (0,67, 2,00)^\top$	$\beta = (2,11, 2,00)^\top$
30	$\beta = (0,17, 2,00)^\top$	$\beta = (1,36, 2,00)^\top$

4.1.1 Algoritmo 1

O algoritmo para gerar os dados do modelo de fração de cura log-simétrico com esquemas de ativação pode ser descrito pelos seguintes passos:

1. Escolher os valores verdadeiros dos parâmetros λ e da pcNC desejada, sabendo que a censura total na amostra é dada por $pc = fc + pcNC$ com fc fixada;
2. Gerar um vetor amostral de covariáveis $X_i = (1, x_i)^\top$ com x_i de tamanho n proveniente de uma distribuição Bernoulli com probabilidade 0,5 para o Cenário 1 e distribuição normal padrão para o Cenário 2;
3. Gerar uma amostra de tamanho n para R_i , assumindo valor constante igual a 1 no caso do esquema de primeira ativação, seguindo uma distribuição uniforme discreta no caso do esquema de ativação aleatória ou uma distribuição geométrica no caso do esquema de ativação geométrica. Em seguida, utilizar o vetor amostral obtido no passo 2 para gerar uma amostra de tamanho n de $M_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_i)$ com $\theta_i = \exp\{X_i^\top \beta\}$.
4. Para cada $i = 1, \dots, n$, gerar uma amostra $\tilde{Y}_j \sim LS(\eta, \phi, g(\cdot))$ com $j = 1, \dots, m_i$. Se $m_i < r_i$, definir $Y_i = 1000 \times 10^{50}$; caso contrário, $Y_i = \tilde{Y}_{(r_i)}$, em que $\tilde{Y}_{(r_i)}$ é a r_i -ésima estatística de ordem da amostra $\tilde{Y}_j$;
5. Encontrar a raiz s da equação $\frac{1}{n} \sum_{i=1}^n F(Y_i) - pc = 0$, sabendo que $F(\cdot)$ é a função de distribuição acumulada (FDA) de uma distribuição uniforme $U(0, s)$;

6. Gerar uma amostra C_i , com $i = 1, \dots, n$ a partir de uma distribuição $U(0, s)$ utilizando o s encontrado no passo 5;
7. Para a censura ser independente do tempo criado, gerar uma nova amostra Y_i utilizando o mesmo procedimento descrito nos passos 3 e 4.
8. Usar a nova amostra, obtida no passo 7, de forma que $T_i = Y_i$ e $\delta_i = 1$ se $Y_i \leq C_i$; caso contrário, $T_i = C_i$ e $\delta_i = 0$.

O desempenho dos estimadores foi avaliado por meio do **VRA**, **DP** e **REQM**, dados por:

$$\begin{aligned} \text{VRA} &= \left| \frac{\text{média}(\hat{\lambda}) - \lambda}{\lambda} \right| \quad (\text{VRA}), \\ \text{DP} &= \sqrt{\frac{\sum_{MC=1}^{10000} (\hat{\lambda}_{MC} - \bar{\hat{\lambda}})^2}{10000 - 1}} \quad (\text{DP}), \\ \text{REQM} &= \sqrt{\frac{1}{10000} \sum_{MC=1}^{10000} (\hat{\lambda}_{MC} - \lambda)^2} \quad (\text{REQM}), \end{aligned}$$

em que MC representa a MC -ésima replica de Monte Carlo sendo utilizadas 10.000 réplicas. Ao longo das diferentes configurações simuladas foi possível observar uma taxa de réplicas que não convergiram, devido a exceder o limite do número máximo de iterações para convergência do processo. No entanto, essa taxa de descarte não ultrapassou 2% em nenhuma das configurações. Em cada réplica, os parâmetros foram estimados por máxima verossimilhança, utilizando a função de verossimilhança marginal através do algoritmo quase-Newton *BFGS* implementado no pacote *maxLik* através do ambiente R ([TEAM et al., 2020](#)) com limite máximo escolhido de 2000 iterações. Os códigos utilizados nesta dissertação podem ser disponibilizados mediante solicitação ao autor.

Os valores iniciais para o processo iterativo de estimação, nos esquemas de primeira ativação e ativação aleatória, foram escolhidos da seguinte forma: para o parâmetro η , utilizou-se a mediana amostral; para ϕ , foi empregada a função quantílica da classe **LS** dada conforme a Propriedade 1 **3.2**. Assim, o valor inicial de ϕ foi determinado por $\phi = \{[\log(\tau_p/\eta)]/Z_\zeta^{(p)}\}^2$, em que $Z_\zeta^{(p)}$ representa o quantil de ordem p da distribuição de latência escolhida. Para o vetor de parâmetros β , as estimativas iniciais foram obtidas por regressão via mínimos quadrados. No caso da distribuição log-t-Student, considerou-se o parâmetro dos graus de liberdade como conhecido e fixo ao longo das simulações.

No esquema de ativação geométrica, um parâmetro adicional foi incorporado à modelagem da fc , o parâmetro ξ . O valor inicial de ξ no processo iterativo foi obtido resolvendo-se a equação de verossimilhança apresentada em (3.7), de forma que

$$\hat{\xi} = \frac{\sum_{i=1}^n v_i}{\sum_{i=1}^n \theta_i F_{T_i}} = \frac{\bar{v}}{\hat{\theta} F_{T_i}(t_i)}.$$

4.1.2 Cenário 1

A seguir, são apresentados os resultados da simulação no Cenário 1, no qual a covariável x_i assume valor 1 para o grupo de tratamento e 0 para o grupo de controle. O parâmetro $\theta_i = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$ depende da covariável binária x_i , com $i = 1, \dots, n$, que representa esses dois grupos. Assume-se que a probabilidade de um indivíduo pertencer ao grupo de tratamento seja de 50% e 50% para o grupo de controle.

As Tabelas 4 e 5, juntamente com a Figura 2, apresentam os resultados das 48 configurações de fc , $pcNC$ e tamanho amostral simuladas no esquema de primeira ativação, $R = 1$, com latência log-normal e log- t -Student com três graus de liberdade. De modo geral, para todos os estimadores e em ambas as distribuições de latência, observam-se reduções no VRA , DP e $REQM$ à medida que o tamanho amostral aumenta, e aumentos nessas métricas com o crescimento da censura, seja ela decorrente do aumento do percentual de censura na fc ou na $pcNC$.

Destaca-se que o estimador do parâmetro associado à mediana apresenta os maiores valores de VRA , DP e $REQM$. Além disso, a distribuição log- t -Student com três graus de liberdade tende a produzir VRA , DP e $REQM$ menores para $\hat{\eta}$ em comparação à log-normal. A Figura 2 também evidencia que, em ambas as distribuições, o impacto no aumento das métricas conforme $pcNC$ aumenta é intensificado à medida que fc cresce, por exemplo, o caso do β_1 para a distribuição log-normal com $n = 500$, o aumento da $pcNC$ de 15% para 45% aumenta o $REQM$ em 27%, 32,5% e 42,6% para $fc = 10\%$, $fc = 20\%$ e $fc = 30\%$, respectivamente.

Tabela 4 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de primeira ativação com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,3270	6,2702	6,0500	0,0689	0,4522	0,4470	0,1330	0,4405	0,4280	0,2390	0,3454	0,3440
		150	0,0536	1,3889	1,3600	0,0204	0,2247	0,2240	0,0306	0,2064	0,2050	0,0222	0,1953	0,1950
		300	0,0192	0,8286	0,8230	0,0055	0,1519	0,1520	0,0078	0,1319	0,1320	0,0381	0,1381	0,1380
		500	0,0157	0,6171	0,6120	0,0068	0,1191	0,1190	0,0107	0,1048	0,1040	0,0025	0,1111	0,1110
20	45	50	1,8465	26,8935	25,3000	0,1236	0,7509	0,7410	0,3502	1,1094	1,0800	0,2002	0,4603	0,4590
		150	0,6347	13,1306	12,7000	0,0639	0,5047	0,5010	0,1294	0,6758	0,6680	0,0304	0,2532	0,2530
		300	0,2593	5,2636	5,1000	0,0451	0,3468	0,3440	0,0980	0,4551	0,4490	0,0304	0,1800	0,1800
		500	0,1465	4,0939	4,0300	0,0296	0,2625	0,2610	0,0583	0,3381	0,3350	0,0005	0,1412	0,1410
30	15	50	0,2667	5,8618	5,7100	0,0720	0,4823	0,4770	0,0964	0,3876	0,3860	0,2182	0,3897	0,3880
		150	0,0571	1,4031	1,3700	0,0242	0,2358	0,2340	0,0661	0,2021	0,2000	0,0223	0,2122	0,2120
		300	0,0220	0,7785	0,7710	0,0052	0,1575	0,1570	0,0182	0,1330	0,1330	0,0362	0,1512	0,1510
		500	0,0119	0,5762	0,5730	0,0066	0,1222	0,1220	0,0071	0,1023	0,1020	0,0069	0,1232	0,1230
45	15	50	1,6633	26,2999	25,0000	0,0809	0,7735	0,7690	0,3261	1,0731	1,0700	0,2339	0,5439	0,5430
		150	0,7577	13,6538	13,1000	0,0800	0,5627	0,5570	0,2860	0,7227	0,7140	0,0094	0,2916	0,2920
		300	0,2176	5,8167	5,7100	0,0105	0,3643	0,3640	0,0557	0,4670	0,4660	0,0192	0,2035	0,2040
		500	0,1356	3,4907	3,4200	0,0253	0,2909	0,2900	0,0757	0,3488	0,3480	0,0019	0,1626	0,1630
15	15	50	0,2356	6,4089	6,3000	0,0307	0,4935	0,4930	0,3813	0,3969	0,3950	0,0058	0,4021	0,4020
		150	0,0454	1,2913	1,2700	0,0173	0,2462	0,2460	0,0963	0,1989	0,1990	0,0231	0,2279	0,2280
		300	0,0265	0,8572	0,8470	0,0151	0,1747	0,1740	0,0929	0,1405	0,1400	0,0056	0,1708	0,1710
		500	0,0124	0,5886	0,5850	0,0050	0,1290	0,1290	0,0135	0,1062	0,1060	0,0191	0,1362	0,1360
45	15	50	2,3342	32,7849	30,6000	0,0495	0,8784	0,8770	1,6184	1,3262	1,3100	0,0400	0,7203	0,7200
		150	1,1645	19,7867	18,9000	0,0919	0,6643	0,6580	1,2070	0,8937	0,8840	0,0117	0,3339	0,3340
		300	0,5420	11,1595	10,8000	0,0585	0,4960	0,4930	0,6669	0,6497	0,6460	0,0270	0,2500	0,2500
		500	0,3660	8,4549	8,2500	0,0497	0,3938	0,3910	0,5778	0,5239	0,5200	0,0465	0,1941	0,1940

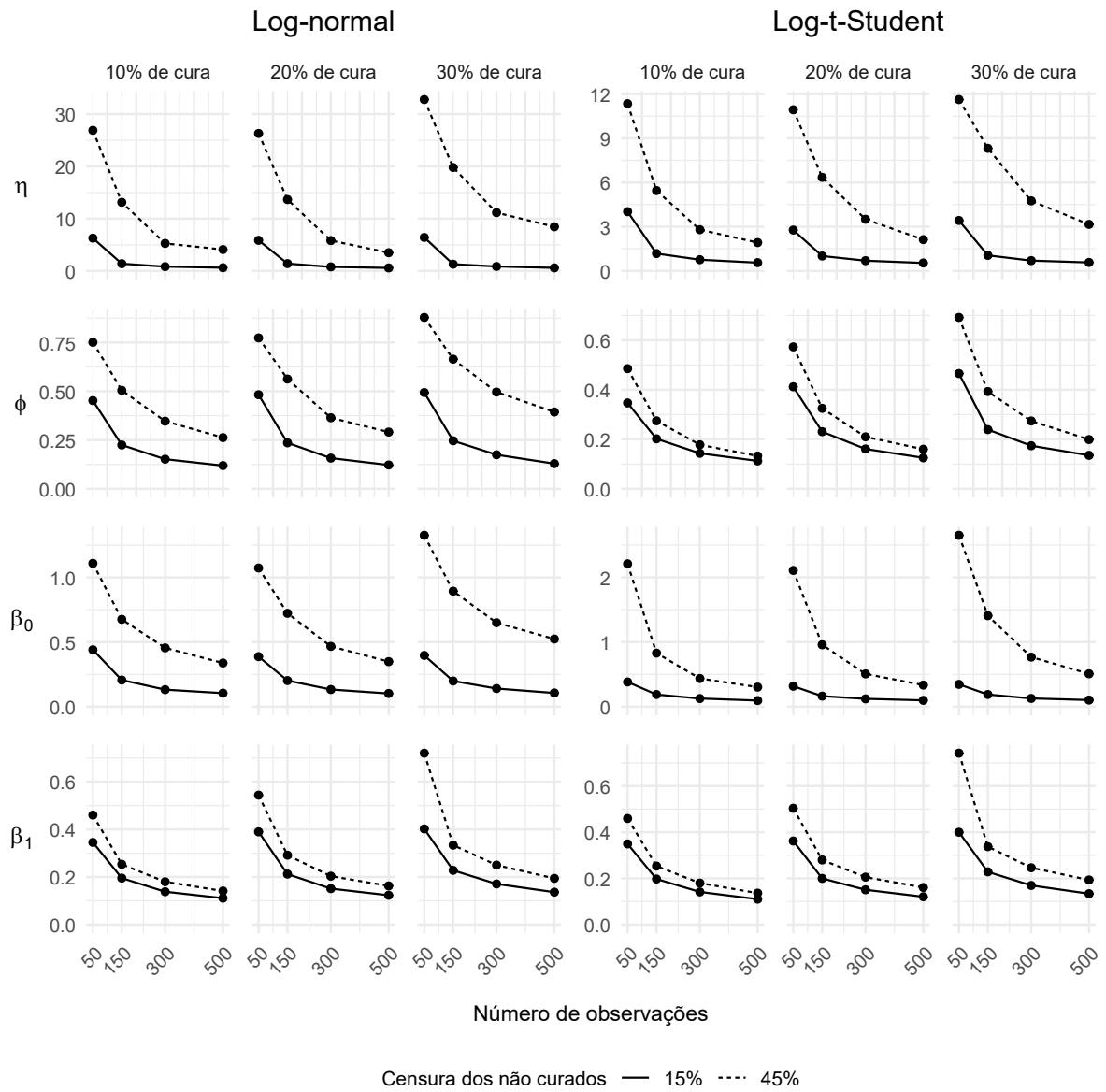
Fonte: o autor

Tabela 5 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de primeira ativação com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.

fc(%)		pcNC(%)	n	η				ϕ				β_0				β_1			
				VRA	REQM	DP		VRA	REQM	DP		VRA	REQM	DP		VRA	REQM	DP	
10	15		50	0,1883	4,0211	3,9100		0,0097	0,3467	0,3470		0,0985	0,3825	0,3750		0,1241	0,3493	0,3490	
			150	0,0465	1,1789	1,1600		0,0038	0,2020	0,2020		0,0321	0,1881	0,1860		0,0530	0,1974	0,1970	
			300	0,0215	0,7625	0,7550		0,0020	0,1435	0,1440		0,0182	0,1264	0,1260		0,0156	0,1414	0,1410	
			500	0,0089	0,5622	0,5600		0,0001	0,1127	0,1130		0,0040	0,0951	0,0951		0,0037	0,1101	0,1100	
	45		50	0,7621	11,3443	10,7000		0,2273	0,4853	0,4290		0,8734	2,2101	2,1100		0,1546	0,4595	0,4590	
			150	0,3116	5,4519	5,2200		0,0873	0,2747	0,2600		0,2385	0,8295	0,8090		0,0968	0,2533	0,2530	
			300	0,1282	2,7993	2,7300		0,0395	0,1780	0,1740		0,0904	0,4362	0,4310		0,0137	0,1798	0,1800	
			500	0,0707	1,9234	1,8900		0,0253	0,1326	0,1300		0,0446	0,3027	0,3010		0,0073	0,1362	0,1360	
	15		50	0,1139	2,7717	2,7100		0,0173	0,4121	0,4120		0,0593	0,3169	0,3160		0,0869	0,3620	0,3620	
			150	0,0222	1,0130	1,0100		0,0020	0,2307	0,2310		0,0035	0,1634	0,1630		0,0287	0,1998	0,2000	
			300	0,0121	0,6964	0,6940		0,0016	0,1610	0,1610		0,0052	0,1214	0,1210		0,0213	0,1507	0,1510	
			500	0,0135	0,5427	0,5380		0,0048	0,1253	0,1250		0,0144	0,0982	0,0980		0,0044	0,1205	0,1200	
20	45		50	0,6452	10,9331	10,4000		0,2062	0,5732	0,5350		1,3539	2,1077	2,0400		0,1224	0,5035	0,5030	
			150	0,3451	6,3475	6,1100		0,0958	0,3250	0,3110		0,4702	0,9579	0,9390		0,0257	0,2797	0,2800	
			300	0,1635	3,5064	3,4100		0,0507	0,2100	0,2040		0,2040	0,5075	0,5010		0,0259	0,2055	0,2060	
			500	0,0749	2,1263	2,0932		0,0237	0,1600	0,1580		0,0794	0,3341	0,3330		0,0198	0,1606	0,1610	
	15		50	0,1037	3,4233	3,3840		0,0208	0,4655	0,4650		0,0372	0,3454	0,3450		0,0489	0,3998	0,4000	
			150	0,0217	1,0569	1,0513		0,0046	0,2391	0,2390		0,0301	0,1884	0,1880		0,0762	0,2283	0,2280	
			300	0,0084	0,7015	0,7003		0,0053	0,1741	0,1740		0,0129	0,1278	0,1280		0,0103	0,1695	0,1700	
			500	0,0084	0,5765	0,5750		0,0024	0,1353	0,1350		0,0289	0,1026	0,1030		0,0227	0,1337	0,1340	
30	45		50	0,7365	11,6261	11,0280		0,2584	0,6921	0,6420		6,6336	2,6508	2,5500		0,0814	0,7421	0,7420	
			150	0,5263	8,3176	7,8900		0,1656	0,3928	0,3560		3,1207	1,4079	1,3700		0,0566	0,3375	0,3370	
			300	0,2311	4,7552	4,6129		0,0843	0,2741	0,2610		1,1278	0,7673	0,7570		0,0124	0,2461	0,2460	
			500	0,1346	3,1600	3,0877		0,0386	0,1990	0,1950		0,5140	0,5091	0,5060		0,0449	0,1932	0,1931	

Fonte: o autor

Figura 2 – **REQM** dos parâmetros no esquema de primeira ativação com latência log-normal e log-t-Student com três graus de liberdade sob diferentes níveis de censura e cura.



Fonte: o autor

Tabela 6 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação aleatória com latência log-normal, sob diferentes níveis de censura e cura.

fc(%) pcNC(%) n		η				ϕ				β_0				β_1			
		VRA	REQM	DP		VRA	REQM	DP		VRA	REQM	DP		VRA	REQM	DP	
10	15	50	0,0103	0,9299	0,9280	0,0264	0,2745	0,2730		0,5990	1,3270	1,2400		1,9600	1,1566	1,1200	
		150	0,0076	0,5545	0,5530	0,0029	0,1702	0,1700		0,1090	0,6854	0,6800		0,7920	0,6515	0,6410	
		300	0,0056	0,3987	0,3980	0,0014	0,1195	0,1200		0,0316	0,2569	0,2560		0,1960	0,3011	0,3000	
		500	0,0018	0,3009	0,3010	0,0004	0,0952	0,0952		0,0210	0,1871	0,1860		0,0898	0,1784	0,1780	
	45	50	0,1186	1,2944	1,1500	0,1686	0,3724	0,3320		0,1414	1,1000	1,0900		1,3682	1,3754	1,3700	
		150	0,0577	0,9070	0,8600	0,0736	0,2347	0,2230		0,1065	0,7465	0,7420		0,9729	0,9797	0,9730	
		300	0,0237	0,6993	0,6890	0,0356	0,1738	0,1700		0,1443	0,6238	0,6140		0,7230	0,7977	0,7920	
		500	0,0152	0,6226	0,6180	0,0227	0,1472	0,1460		0,1219	0,5313	0,5230		0,9220	0,6837	0,6720	
	15	50	0,0420	1,2330	1,2200	0,0159	0,3559	0,3560		0,3381	0,6921	0,6790		0,5350	0,9409	0,9240	
		150	0,0095	0,6277	0,6260	0,0004	0,1845	0,1840		0,0535	0,2133	0,2120		0,6280	0,3412	0,3410	
		300	0,0074	0,4627	0,4610	0,0010	0,1317	0,1320		0,0040	0,1399	0,1400		0,4630	0,2193	0,2190	
		500	0,0001	0,3328	0,3330	0,0033	0,0982	0,0982		0,0034	0,1062	0,1060		0,3330	0,2148	0,2140	
	45	50	0,0824	1,8106	1,7600	0,1497	0,4582	0,4330		0,2225	1,1263	1,1200		1,4459	1,4589	1,4500	
		150	0,0355	1,2594	1,2500	0,0683	0,2874	0,2790		0,2686	0,7781	0,7710		1,2600	0,8751	0,8710	
		300	0,0012	1,0794	1,0800	0,0310	0,2287	0,2270		0,2955	0,6551	0,6440		1,0800	0,7294	0,7230	
		500	0,0027	0,9222	0,9220	0,0149	0,1827	0,1820		0,1848	0,4888	0,4830		0,9220	0,5302	0,5240	
	15	50	0,0629	1,7204	1,6900	0,0095	0,4170	0,4170		0,9729	0,8611	0,8540		1,7200	0,8581	0,8500	
		150	0,0138	0,7591	0,7560	0,0019	0,2119	0,2120		0,1562	0,2106	0,2100		0,7590	0,3197	0,3200	
		300	0,0082	0,5250	0,5240	0,0045	0,1515	0,1510		0,0384	0,1423	0,1420		0,5250	0,2057	0,2060	
		500	0,0057	0,3882	0,3870	0,0012	0,1167	0,1170		0,0099	0,1106	0,1110		0,3880	0,1596	0,1600	
	45	50	0,0684	2,5317	2,5100	0,1720	0,5704	0,5440		1,1369	1,9174	1,9100		2,2146	1,9983	1,9700	
		150	0,0066	1,8759	1,8800	0,0598	0,3680	0,3630		1,5235	0,9827	0,9680		1,1000	1,0341	1,0200	
		300	0,0212	1,6087	1,6100	0,0247	0,2974	0,2960		1,3890	0,8173	0,8030		1,6100	0,7717	0,7570	
		500	0,0568	1,6107	1,5900	0,0048	0,2650	0,2650		2,8876	1,0225	0,9720		0,0651	0,6624	0,6620	

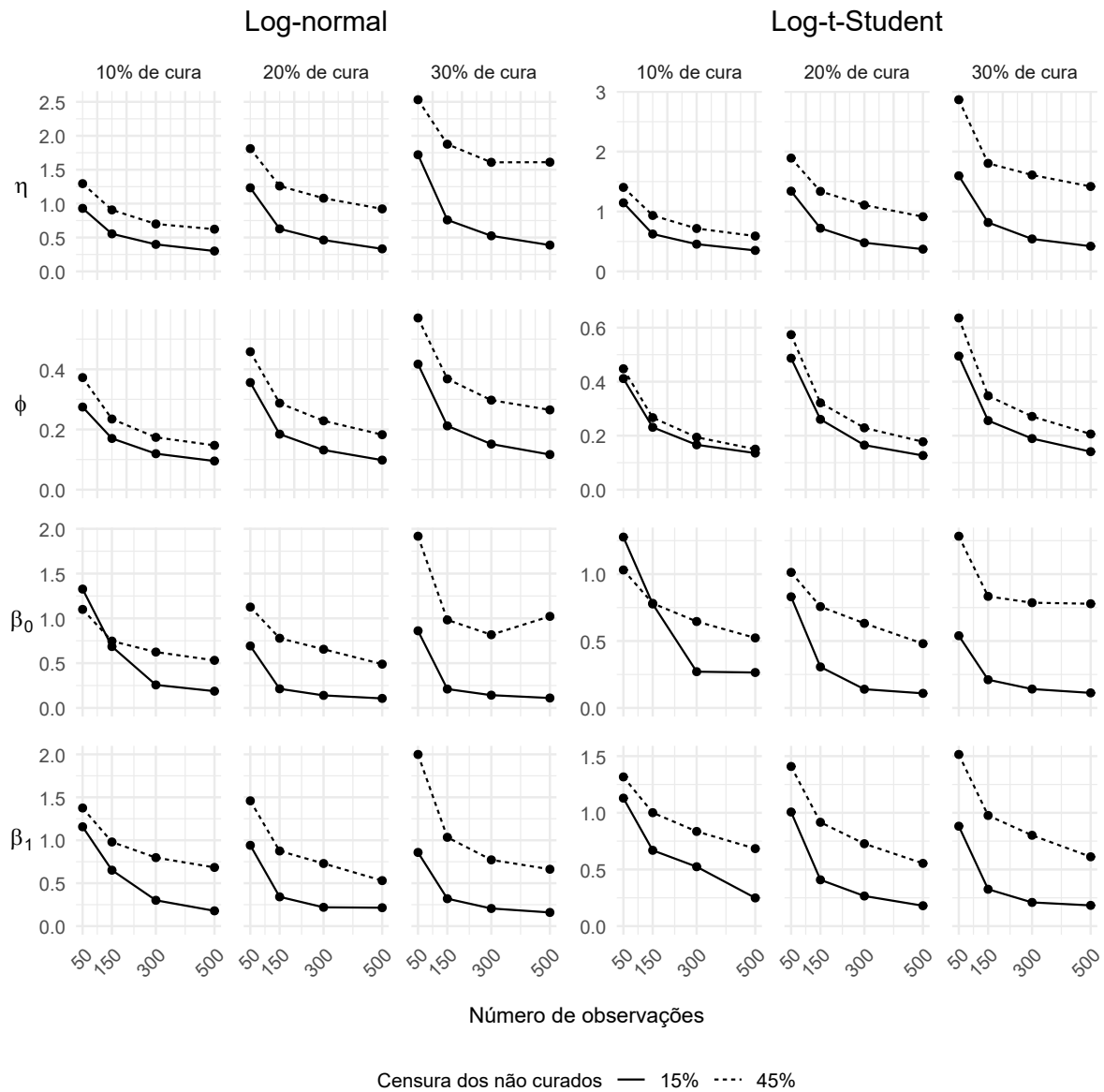
Fonte: o autor

Tabela 7 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação aleatória com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,0186	1,1453	1,1400	0,0218	0,4115	0,4110	0,6142	1,2768	1,1900	1,1040	1,5826	1,1000
		150	0,0083	0,6243	0,6230	0,0094	0,2310	0,2310	0,2136	0,7767	0,7590	0,6537	0,9439	0,6540
		300	0,0100	0,4547	0,4520	0,0147	0,1661	0,1650	0,0486	0,2708	0,2680	0,5065	0,9029	0,5060
		500	0,0081	0,3508	0,3480	0,0078	0,1355	0,1350	0,0469	0,2647	0,2620	0,2476	0,2490	0,2480
	45	50	0,0882	1,4032	1,3300	0,0916	0,4478	0,4380	0,2141	1,0310	1,0200	1,3065	1,0811	1,3100
		150	0,0418	0,9342	0,9110	0,0474	0,2667	0,2620	0,2158	0,7816	0,7640	0,4285	1,0010	0,9990
		300	0,0146	0,7166	0,7130	0,0244	0,1942	0,1930	0,1899	0,6459	0,6290	0,8288	0,8355	0,8290
		500	0,0030	0,5905	0,5900	0,0104	0,1500	0,1500	0,1514	0,5232	0,5100	0,6769	0,6841	0,6770
	15	50	0,0452	1,3403	1,3200	0,0675	0,4869	0,4820	0,4387	0,8306	0,8120	0,9847	1,3984	0,9850
		150	0,0134	0,7212	0,7180	0,0180	0,2604	0,2600	0,0638	0,3065	0,3050	0,4048	0,4088	0,4050
		300	0,0006	0,4791	0,4790	0,0072	0,1650	0,1650	0,0015	0,1398	0,1400	0,2657	0,2660	0,2660
		500	0,0028	0,3716	0,3710	0,0032	0,1265	0,1260	0,0086	0,1089	0,1090	0,1802	0,1802	0,1800
20	45	50	0,0646	1,8922	1,8600	0,0548	0,5741	0,5720	0,2133	1,0131	1,0100	1,4011	0,9732	1,4000
		150	0,0001	1,3369	1,3400	0,0291	0,3215	0,3200	0,3793	0,7570	0,7420	0,9085	0,9163	0,9080
		300	0,0207	1,1079	1,1000	0,0182	0,2290	0,2280	0,3938	0,6322	0,6120	0,7186	0,7272	0,7190
		500	0,0168	0,9137	0,9100	0,0049	0,1773	0,1770	0,2386	0,4803	0,4710	0,5469	0,5538	0,5470
	15	50	0,0533	1,5963	1,5700	0,0753	0,4947	0,4890	0,6703	0,5396	0,5340	0,8756	0,6954	0,8760
		150	0,0107	0,8169	0,8150	0,0090	0,2558	0,2560	0,0191	0,2106	0,2110	0,1576	0,3256	0,3250
		300	0,0086	0,5424	0,5410	0,0044	0,1893	0,1890	0,0265	0,1410	0,1410	0,0081	0,2095	0,2100
		500	0,0078	0,4205	0,4190	0,0026	0,1408	0,1410	0,0018	0,1123	0,1120	0,1832	0,1833	0,1830
30	45	50	0,0838	2,8664	2,8400	0,0887	0,6359	0,6300	0,3119	1,2833	1,2800	1,5061	1,5145	1,5100
		150	0,0044	1,8044	1,8000	0,0469	0,3476	0,3440	1,1029	0,8338	0,8250	0,7431	0,9765	0,9700
		300	0,0323	1,6101	1,6000	0,0200	0,2712	0,2700	1,4680	0,7862	0,7700	1,0075	0,8016	0,7870
		500	0,0443	1,4183	1,4000	0,0072	0,2061	0,2060	1,6703	0,7785	0,7560	0,2598	0,6118	0,6110

Fonte: o autor

Figura 3 – **REQM** dos parâmetros no esquema de ativação aleatória com latência log-normal e log-t-Student com três graus de liberdade sob diferentes níveis de censura e cura.

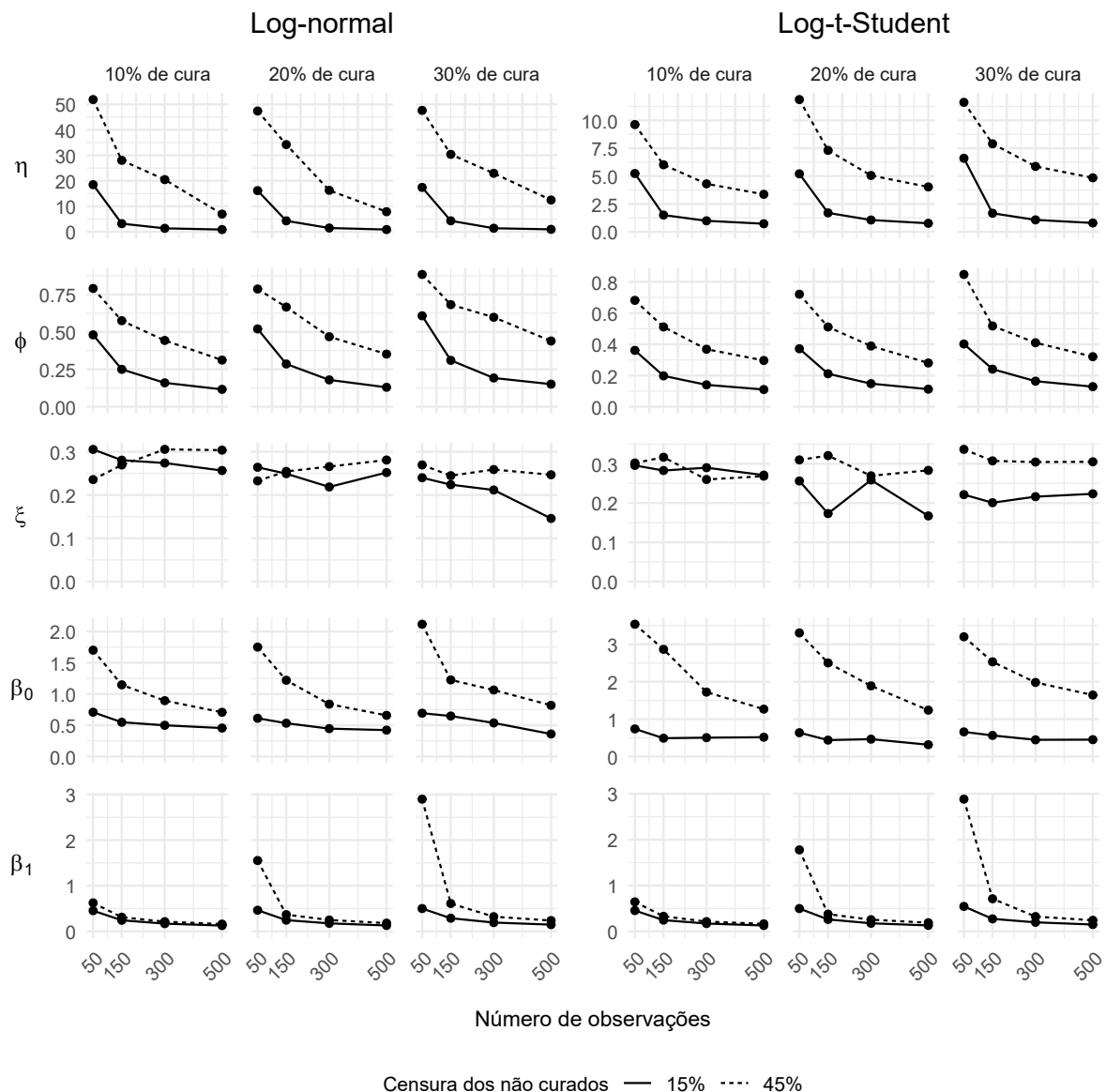


Fonte: o autor

De forma similar com o observado nos resultados do esquema de primeira ativação, as Tabelas 6 e 7 juntamente com a Figura 3 mostram que para todos os estimadores e em ambas as distribuições de latência, o **VRA**, o **DP** e o **REQM** do esquema de ativação aleatória, $R \sim U(1, M)$, tendem a diminuir com o aumento do tamanho amostral e a crescer com o aumento percentual de censura, seja ela causada pelo aumento da **fc** ou da **pcNC**. Esse efeito, no entanto, é atenuado em amostras maiores. Em situações onde os **VRA** já se aproximam de zero, flutuações aleatórias podem gerar pequenos aumentos nas métricas mesmo com o crescimento da amostra.

Observa-se também que, diferentemente do esquema de primeira ativação, no qual o estimador $\hat{\eta}$ apresenta métricas elevadas, neste caso as métricas permanecem bastante próximas entre os diferentes estimadores em ambas as distribuições. O mesmo pode ser constatado ao comparar as métricas entre as distribuições. Também nota-se na Figura 3 que assim como no esquema de primeira ativação, o crescimento da fc eleva o efeito do aumento da $pcNC$.

Figura 4 – **REQM** dos parâmetros no esquema de ativação geométrica com latência log-normal e log-t-Student com três graus de liberdade sob diferentes níveis de censura e cura.



Fonte: o autor

Por fim, os resultados das simulações de Monte Carlo no esquema de ativação geométrica, $R \sim \text{Geom}(\xi)$, com latência log-normal e log-t-Student com três graus de liberdade são apresentados nas Tabelas 9, ?? e na Figura 4. De maneira geral, exceto para o estimador $\hat{\xi}$, os

Tabela 8 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			ξ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,8249	18,4778	18,0126	0,0647	0,4803	0,4759	0,4590	0,3053	0,2014	0,1456	0,7086	0,6864	0,0511	0,4536	0,4420
		150	0,1618	3,2003	3,0965	0,0308	0,2497	0,2478	0,3310	0,2805	0,2265	0,1292	0,5480	0,5252	0,0179	0,2434	0,2407
		300	0,0542	1,3641	1,3370	0,0089	0,1596	0,1593	0,2790	0,2739	0,2357	0,1193	0,4992	0,4779	0,0095	0,1697	0,1687
		500	0,0329	0,8830	0,8676	0,0088	0,1160	0,1157	0,2696	0,2566	0,2183	0,1259	0,4551	0,4289	0,0067	0,1274	0,1267
	45	50	4,2363	51,8692	47,3496	0,1603	0,7905	0,7741	0,2566	0,2357	0,1977	0,2615	1,7017	1,6721	0,0608	0,6252	0,6133
		150	1,8199	28,0238	26,5068	0,1220	0,5749	0,5618	0,3452	0,2694	0,2069	0,0538	1,1470	1,1452	0,0157	0,3107	0,3092
		300	1,0922	20,4721	19,7313	0,1052	0,4429	0,4303	0,4586	0,3055	0,2020	0,0742	0,8934	0,8890	0,0104	0,2132	0,2122
		500	0,3687	6,9428	6,6939	0,0518	0,3116	0,3073	0,4333	0,3037	0,2128	0,1598	0,7076	0,6807	0,0075	0,1600	0,1593
	15	50	0,7592	16,1611	15,7097	0,0914	0,5199	0,5118	0,2616	0,2641	0,2295	0,0488	0,6119	0,6110	0,0551	0,4637	0,4504
		150	0,1689	4,3042	4,2208	0,0322	0,2856	0,2838	0,1805	0,2493	0,2324	0,0550	0,5328	0,5315	0,0164	0,2467	0,2445
		300	0,0610	1,4870	1,4554	0,0133	0,1792	0,1787	0,1375	0,2187	0,2076	0,0474	0,4459	0,4448	0,0067	0,1759	0,1754
		500	0,0229	0,8897	0,8824	0,0032	0,1297	0,1297	0,2909	0,2519	0,2057	0,2883	0,4225	0,3758	0,0047	0,1297	0,1293
	45	50	3,5426	47,3821	43,9489	0,0953	0,7869	0,7812	0,0456	0,2327	0,2316	0,3125	1,7509	1,7384	0,1620	1,5523	1,5182
		150	2,4191	34,2082	32,0001	0,1661	0,6661	0,6451	0,2419	0,2545	0,2239	0,3092	1,2200	1,2023	0,0202	0,3668	0,3646
		300	0,9425	16,2621	15,5652	0,0849	0,4679	0,4601	0,2600	0,2657	0,2318	0,0270	0,8380	0,8378	0,0131	0,2493	0,2480
		500	0,4475	7,8897	7,5661	0,0576	0,3516	0,3468	0,3084	0,2809	0,2348	0,0951	0,6584	0,6554	0,0069	0,1810	0,1804
	15	50	0,9097	17,4193	16,8158	0,1186	0,6077	0,5961	0,0367	0,2399	0,2392	1,1495	0,6929	0,6648	0,0532	0,5002	0,4887
		150	0,1725	4,3174	4,2306	0,0321	0,3100	0,3084	0,2483	0,2239	0,1863	2,4187	0,6479	0,5007	0,0215	0,2883	0,2851
		300	0,0618	1,3939	1,3594	0,0169	0,1919	0,1911	0,0208	0,2116	0,2113	0,5136	0,5369	0,5298	0,0119	0,1936	0,1921
		500	0,0361	0,9688	0,9518	0,0132	0,1508	0,1502	0,0865	0,1459	0,1393	0,8555	0,3604	0,3297	0,0043	0,1480	0,1477
	45	50	3,7885	47,6179	43,6902	0,0661	0,8841	0,8817	0,2459	0,2695	0,2398	1,3864	2,1165	2,1034	0,4360	2,8968	2,7625
		150	2,1450	30,3811	28,4265	0,1111	0,6824	0,6734	0,0282	0,2447	0,2443	2,1966	1,2254	1,1672	0,0512	0,6107	0,6021
		300	1,5683	22,9469	21,5666	0,1407	0,5977	0,5809	0,0394	0,2589	0,2582	2,1018	1,0637	1,0019	0,0225	0,3226	0,3195
		500	0,6805	12,4738	12,0013	0,0676	0,4392	0,4340	0,0075	0,2468	0,2468	1,5393	0,8194	0,7765	0,0089	0,2373	0,2366

Fonte: o autor

Tabela 9 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			ξ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,2199	5,2334	5,1168	0,0946	0,3609	0,3483	0,4004	0,2957	0,2176	0,1457	0,7374	0,7168	0,0517	0,4546	0,4427
		150	0,0405	1,5034	1,4898	0,0300	0,1976	0,1953	0,4050	0,2827	0,1973	0,2373	0,4894	0,3998	0,0201	0,2455	0,2422
		300	0,0209	0,9866	0,9811	0,0106	0,1401	0,1397	0,2899	0,2899	0,2511	0,1333	0,5037	0,4781	0,0108	0,1712	0,1699
		500	0,0090	0,7241	0,7227	0,0090	0,1102	0,1098	0,2448	0,2710	0,2418	0,1068	0,5157	0,4998	0,0055	0,1285	0,1281
20	45	50	0,5729	9,6250	9,1893	0,2748	0,6815	0,6237	0,2888	0,3018	0,2650	1,1621	3,5369	3,2555	0,0626	0,6426	0,6303
		150	0,3885	6,0187	5,6969	0,2125	0,5109	0,4646	0,4423	0,3165	0,2264	0,7008	2,8661	2,7423	0,0275	0,3286	0,3240
		300	0,2202	4,3072	4,1643	0,1123	0,3684	0,3509	0,1975	0,2599	0,2404	0,3114	1,7219	1,6817	0,0131	0,2138	0,2122
		500	0,1618	3,3638	3,2652	0,0850	0,2963	0,2839	0,2587	0,2684	0,2352	0,1047	1,2658	1,2597	0,0054	0,1642	0,1639
30	15	50	0,2473	5,2072	5,0585	0,0695	0,3714	0,3649	0,1534	0,2561	0,2444	0,0331	0,6365	0,6361	0,0614	0,4980	0,4826
		150	0,0723	1,6976	1,6588	0,0289	0,2111	0,2091	0,0522	0,1732	0,1712	0,2145	0,4381	0,4154	0,0240	0,2607	0,2563
		300	0,0265	1,0594	1,0511	0,0148	0,1478	0,1470	0,2044	0,2586	0,2375	0,1510	0,4634	0,4529	0,0076	0,1765	0,1759
		500	0,0137	0,7628	0,7597	0,0071	0,1123	0,1120	0,0810	0,1672	0,1623	0,0526	0,3148	0,3129	0,0026	0,1302	0,1301
45	15	50	0,6718	11,8716	11,3871	0,2585	0,7207	0,6728	0,0530	0,3098	0,3087	1,9156	3,3047	3,0614	0,2049	1,7767	1,7289
		150	0,4950	7,3107	6,8793	0,2333	0,5107	0,4544	0,2868	0,3207	0,2868	1,3205	2,5028	2,3511	0,0307	0,3761	0,3711
		300	0,3114	5,0582	4,8129	0,1441	0,3889	0,3612	0,0723	0,2693	0,2668	1,2564	1,8890	1,7035	0,0075	0,2557	0,2553
		500	0,1959	4,0293	3,9087	0,0857	0,2801	0,2667	0,1922	0,2832	0,2664	0,2969	1,2388	1,2237	0,0021	0,1888	0,1888
15	15	50	0,3546	6,5995	6,3572	0,0629	0,4018	0,3968	0,0692	0,2213	0,2186	1,8016	0,6577	0,6146	0,0721	0,5444	0,5250
		150	0,0633	1,6661	1,6359	0,0089	0,2406	0,2405	0,2984	0,2008	0,1344	3,1791	0,5633	0,3828	0,0165	0,2729	0,2709
		300	0,0336	1,0706	1,0574	0,0048	0,1636	0,1635	0,0368	0,2162	0,2154	0,9141	0,4455	0,4294	0,0145	0,1975	0,1954
		500	0,0203	0,7918	0,7853	0,0023	0,1284	0,1284	0,0728	0,2234	0,2204	1,2723	0,4502	0,4187	0,0073	0,1495	0,1488
45	15	50	0,6209	11,6217	11,1999	0,2890	0,8474	0,7967	0,1546	0,3366	0,3276	8,6003	3,2022	3,0009	0,4597	2,8830	2,7326
		150	0,4995	7,9024	7,4978	0,2361	0,5173	0,4603	0,0413	0,3073	0,3066	8,6155	2,5318	2,2707	0,0517	0,7117	0,7042
		300	0,3499	5,8721	5,6056	0,1603	0,4098	0,3772	0,3030	0,3044	0,2640	8,8898	1,9803	1,6082	0,0174	0,3231	0,3212
		500	0,2530	4,8447	4,6769	0,1113	0,3206	0,3007	0,0300	0,3050	0,3046	4,5469	1,6440	1,5341	0,0157	0,2422	0,2401

Fonte: o autor

resultados mostram que, para todos os estimadores e em ambas as distribuições de latência, o **VRA**, o **DP** e o **REQM** diminuem com o aumento da amostra e aumentam com o crescimento da censura, seja esta oriunda da **fc** ou da **pcNC**. Como nos demais esquemas, esse efeito é menos expressivo em amostras grandes.

O estimador do parâmetro η , associado à mediana, volta a apresentar valores de **DP** e **REQM** altos, destoando-se dos demais estimadores, ainda mais do que no caso do esquema de primeira ativação, sendo que a distribuição log- t -Student com três graus de liberdade continua resultando em métricas inferiores à log-normal. Já o estimador $\hat{\xi}$ não apresenta uma melhora consistente das métricas com o aumento da amostra, mostrando grandes oscilações nas métricas de **VRA**, **DP** e **REQM**. A Figura 4, assim como os esquemas anteriores, também evidencia que, para ambas as distribuições, o aumento da **fc** intensifica o impacto do crescimento da **pcNC**, com exceção do estimador de ξ , onde as oscilações dificultam a detecção desse padrão.

4.1.3 Cenário 2

A seguir, são apresentados os resultados da simulação no Cenário 2 com covariável x_i contínua, em que o parâmetro $\theta_i = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$ é associado a uma covariável contínua com distribuição normal padrão, ou seja, média zero e desvio padrão igual a um, representando, por exemplo, a variação no peso corporal após determinado tratamento.

De forma geral, as Tabelas 10 e 11, e na Figura 5, referentes ao Cenário 2 com esquema de primeira ativação, mostram que as métricas de **VRA**, **DP** e **REQM** diminuem conforme o tamanho amostral aumenta e aumentam conforme o percentual da censura aumenta. Nota-se ainda que o estimador do parâmetro associado a mediana apresenta os maiores valores dessas métricas, especialmente o **DP** e o **REQM**. Esse comportamento é especialmente evidente no estimador $\hat{\eta}$ da latência log-normal, que apresenta métricas consideravelmente maiores que a distribuição log- t -Student com três graus de liberdade. A Figura 5 também evidencia que a medida que a **fc** aumenta, o impacto do acréscimo na **pcNC** se intensifica em ambas as distribuições.

Tabela 10 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de primeira ativação com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,8550	17,1328	16,5915	0,0721	0,4673	0,4617	0,1560	0,7883	0,7521	0,0619	0,2448	0,2342
		150	0,1389	2,6289	2,5356	0,0239	0,2323	0,2310	0,0485	0,3346	0,3265	0,0243	0,1358	0,1329
		300	0,0564	1,4709	1,4437	0,0060	0,1614	0,1613	0,0197	0,2286	0,2266	0,0122	0,0923	0,0912
		500	0,0274	0,9706	0,9610	0,0038	0,1206	0,1206	0,0095	0,1690	0,1684	0,0040	0,0691	0,0689
	45	50	3,8700	49,4264	45,4819	0,1250	0,7393	0,7287	0,3410	1,8471	1,7739	0,0599	0,2866	0,2783
		150	1,5518	23,8985	22,6050	0,1093	0,5237	0,5122	0,2029	1,1875	1,1473	0,0286	0,1689	0,1657
		300	0,6383	12,1221	11,6949	0,0468	0,3872	0,3844	0,0917	0,8260	0,8144	0,0132	0,1094	0,1083
		500	0,3263	6,9455	6,7515	0,0286	0,2937	0,2923	0,0531	0,6473	0,6424	0,0047	0,0847	0,0845
	15	50	0,5593	13,1802	12,8808	0,0533	0,4586	0,4555	0,1725	0,5760	0,5566	0,0568	0,2496	0,2409
		150	0,0940	2,1283	2,0759	0,0136	0,2424	0,2420	0,0434	0,2577	0,2549	0,0235	0,1369	0,1342
		300	0,0521	1,2411	1,2135	0,0110	0,1602	0,1598	0,0315	0,1760	0,1739	0,0115	0,0951	0,0942
		500	0,0258	0,8401	0,8301	0,0042	0,1244	0,1243	0,0200	0,1338	0,1327	0,0047	0,0719	0,0717
	45	50	3,8853	48,4040	44,3369	0,1150	0,8074	0,7992	0,4892	1,8065	1,7569	0,0794	0,3460	0,3337
		150	1,6886	26,8858	25,5269	0,1079	0,5804	0,5703	0,3206	1,1757	1,1430	0,0268	0,1848	0,1822
		300	0,7919	14,1464	13,5817	0,0829	0,4375	0,4296	0,2092	0,8360	0,8164	0,0142	0,1222	0,1211
		500	0,4894	11,6460	11,3866	0,0480	0,3516	0,3484	0,1324	0,6714	0,6617	0,0047	0,0941	0,0939
	15	50	0,5105	9,3089	8,9526	0,0452	0,5095	0,5075	0,3217	0,5372	0,5224	0,0609	0,2694	0,2601
		150	0,0922	2,2360	2,1881	0,0120	0,2688	0,2685	0,0615	0,2604	0,2593	0,0222	0,1479	0,1456
		300	0,0548	1,3722	1,3446	0,0147	0,1900	0,1895	0,0544	0,1829	0,1816	0,0160	0,1032	0,1015
		500	0,0341	0,9484	0,9330	0,0094	0,1447	0,1444	0,0305	0,1360	0,1355	0,0083	0,0784	0,0778
30	45	50	3,5094	42,8159	39,0570	0,0444	0,8445	0,8433	0,7420	1,9614	1,9400	0,0983	0,4372	0,4224
		150	2,2285	31,0344	28,9666	0,1057	0,6859	0,6777	0,7287	1,3593	1,3293	0,0279	0,2107	0,2083
		300	1,0952	17,8970	17,0395	0,0914	0,5167	0,5086	0,5050	0,9670	0,9467	0,0224	0,1465	0,1442
		500	0,7270	12,6249	12,0909	0,0794	0,4344	0,4271	0,4110	0,8091	0,7931	0,0081	0,1062	0,1058

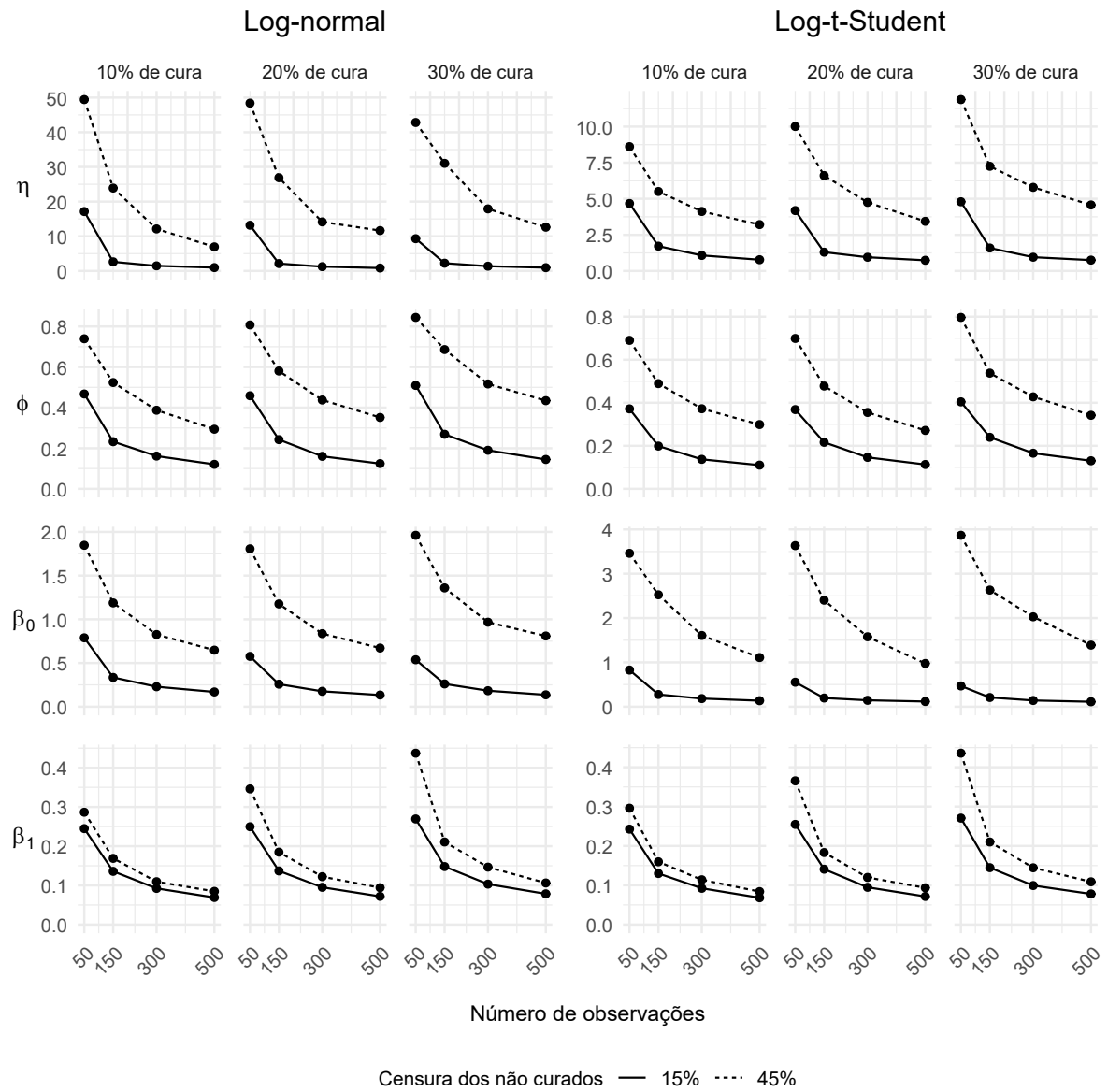
Fonte: o autor

Tabela 11 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de primeira ativação com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.

f _c (%)		pcNC(%)		n		η			ϕ			β_0			β_1		
						VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15		50	0,1980	4,6721	4,5700	0,0944	0,3716	0,3600	0,1080	0,8264	0,8100	0,0504	0,2427	0,2360		
				0,0458	1,7175	1,7000	0,0327	0,1988	0,1960	0,0233	0,2756	0,2730	0,0169	0,1299	0,1280		
				0,0252	1,0761	1,0700	0,0197	0,1371	0,1360	0,0142	0,1833	0,1820	0,0102	0,0923	0,0916		
				0,0113	0,7773	0,7750	0,0103	0,1102	0,1100	0,0064	0,1354	0,1350	0,0042	0,0682	0,0680		
	45		50	0,3820	8,6090	8,4000	0,2370	0,6901	0,6480	0,8230	3,4599	3,2300	0,0663	0,2959	0,2860		
				0,2970	5,5032	5,3000	0,1700	0,4890	0,4580	0,5360	2,5245	2,3900	0,0214	0,1596	0,1580		
				0,2020	4,1211	4,0000	0,1100	0,3722	0,3560	0,2640	1,6066	1,5600	0,0115	0,1139	0,1130		
				0,1500	3,2124	3,1200	0,0829	0,2988	0,2870	0,1670	1,1088	1,0800	0,0025	0,0838	0,0838		
	15		50	0,1430	4,1832	4,1200	0,0840	0,3683	0,3590	0,0795	0,5525	0,5480	0,0596	0,2547	0,2450		
				0,0319	1,3007	1,2900	0,0194	0,2163	0,2150	0,0129	0,1956	0,1950	0,0163	0,1408	0,1400		
				0,0256	0,9495	0,9410	0,0075	0,1463	0,1460	0,0102	0,1444	0,1440	0,0096	0,0948	0,0942		
				0,0064	0,7389	0,7380	0,0070	0,1131	0,1130	0,0000	0,1169	0,1170	0,0051	0,0715	0,0712		
20	45		50	0,4800	10,0082	9,7200	0,2820	0,6989	0,6390	1,5200	3,6333	3,3900	0,0797	0,3655	0,3540		
				0,3970	6,6075	6,3000	0,1820	0,4777	0,4420	0,8780	2,4027	2,2800	0,0209	0,1831	0,1820		
				0,2510	4,7496	4,5800	0,1230	0,3550	0,3330	0,4420	1,5775	1,5300	0,0163	0,1199	0,1180		
				0,1600	3,4378	3,3400	0,0857	0,2715	0,2580	0,2340	0,9748	0,9540	0,0066	0,0935	0,0932		
	15		50	0,2080	4,7912	4,6800	0,0753	0,4042	0,3970	0,1500	0,4678	0,4640	0,0701	0,2707	0,2580		
				0,0572	1,5901	1,5600	0,0109	0,2395	0,2390	0,0531	0,2077	0,2070	0,0132	0,1447	0,1440		
				0,0176	0,9526	0,9490	0,0092	0,1656	0,1650	0,0011	0,1405	0,1400	0,0132	0,0993	0,0981		
				0,0147	0,7512	0,7480	0,0064	0,1304	0,1300	0,0135	0,1112	0,1110	0,0067	0,0779	0,0776		
30	45		50	0,5620	11,8714	11,5000	0,2990	0,7967	0,7380	3,2000	3,8665	3,6600	0,1120	0,4358	0,4170		
				0,4330	7,2487	6,9200	0,2000	0,5377	0,4990	2,1700	2,6308	2,4900	0,0211	0,2100	0,2090		
				0,3590	5,7892	5,5000	0,1740	0,4276	0,3900	1,5700	2,0277	1,9300	0,0197	0,1444	0,1430		
				0,2600	4,5719	4,3800	0,1260	0,3419	0,3180	0,9430	1,3892	1,3400	0,0094	0,1088	0,1080		

Fonte: o autor

Figura 5 – **REQM** dos parâmetros no esquema de primeira ativação com latência log-normal e log-t-Student com três graus de liberdade sob diferentes níveis de censura e cura.



Fonte: o autor

Tabela 12 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação aleatória com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)		pcNC(%)	n	η			ϕ			β_0			β_1		
				VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15		50	0,0038	0,8126	0,8120	0,0414	0,2573	0,2540	3,1400	14,5549	13,8000	2,6200	9,0454	8,5300
			150	0,0006	0,4966	0,4970	0,0116	0,1459	0,1450	0,3080	3,8047	3,7800	0,2890	2,3686	2,3500
			300	0,0015	0,3496	0,3500	0,0068	0,1072	0,1070	0,1000	0,5688	0,5480	0,1090	0,4550	0,4370
			500	0,0011	0,2628	0,2630	0,0026	0,0800	0,0799	0,0398	0,3008	0,2950	0,0407	0,2536	0,2490
	45		50	0,1118	1,2484	1,1200	0,1576	0,3689	0,3340	4,1804	16,2668	15,0000	2,8113	12,5185	12,1000
			150	0,0379	0,7880	0,7650	0,0511	0,2210	0,2150	3,1684	20,8574	20,3000	2,0534	11,9634	11,7000
			300	0,0230	0,5588	0,5470	0,0324	0,1601	0,1570	0,6230	4,5895	4,4900	0,5247	2,7034	2,6400
			500	0,0117	0,4166	0,4120	0,0166	0,1142	0,1130	0,2881	2,5477	2,5100	0,2530	1,4561	1,4300
	15		50	0,0132	1,0126	1,0100	0,0320	0,2868	0,2850	1,4176	5,9399	5,8100	0,9361	4,5128	4,3800
			150	0,0001	0,5439	0,5440	0,0118	0,1683	0,1680	0,1294	0,6454	0,6360	0,1097	0,6222	0,6090
			300	0,0046	0,3990	0,3980	0,0014	0,1187	0,1190	0,0393	0,2237	0,2210	0,0372	0,2691	0,2660
			500	0,0023	0,2941	0,2940	0,0059	0,0871	0,0869	0,0263	0,1648	0,1630	0,0249	0,1923	0,1900
20	45		50	0,0983	1,6353	1,5600	0,1533	0,4271	0,3990	4,5311	11,0891	10,4000	2,7761	10,0807	9,5600
			150	0,0381	0,9735	0,9550	0,0584	0,2492	0,2420	3,0269	11,4485	11,1000	1,7686	8,1642	7,9100
			300	0,0227	0,7122	0,7030	0,0316	0,1778	0,1750	0,3549	1,9953	1,9700	0,2581	1,4767	1,4500
			500	0,0126	0,5372	0,5340	0,0203	0,1357	0,1340	0,1263	0,8448	0,8380	0,1097	0,6912	0,6800
	15		50	0,0122	1,1873	1,1900	0,0273	0,3290	0,3280	1,9409	4,0146	3,9400	0,8232	3,6432	3,5200
			150	0,0029	0,6347	0,6350	0,0138	0,1921	0,1920	0,1184	0,2829	0,2790	0,0786	0,3964	0,3860
			300	0,0054	0,4704	0,4700	0,0064	0,1378	0,1380	0,0592	0,1760	0,1740	0,0446	0,2631	0,2580
			500	0,0009	0,3399	0,3400	0,0036	0,1022	0,1020	0,0307	0,1282	0,1280	0,0195	0,1748	0,1730
30	45		50	0,1002	3,1005	3,0600	0,1663	0,6011	0,5780	8,4806	10,0342	9,4700	3,2534	10,2712	9,5700
			150	0,0456	1,4114	1,3900	0,0763	0,3335	0,3250	5,4312	9,7967	9,5700	1,4537	7,5142	7,3300
			300	0,0311	1,0396	1,0300	0,0473	0,2354	0,2310	0,6873	1,9918	1,9700	0,2997	1,7274	1,6900
			500	0,0118	0,7680	0,7660	0,0205	0,1745	0,1730	0,2921	1,0851	1,0800	0,1269	0,8683	0,8560

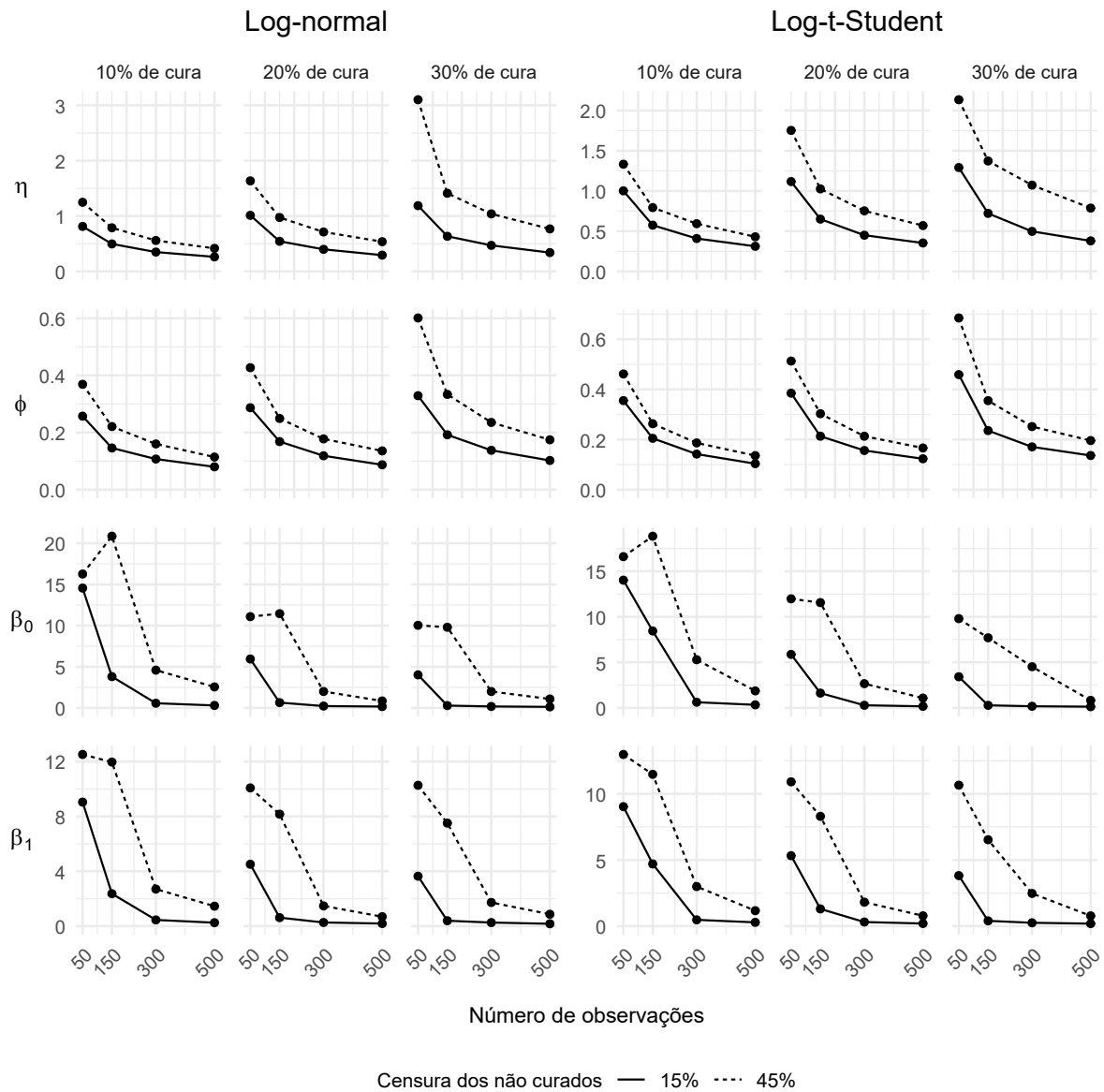
Fonte: o autor

Tabela 13 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação aleatória com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0,0095	1,0019	1,0000	0,0245	0,3552	0,3540	3,2500	14,0209	13,1000	2,7500	9,0275	8,4500
		150	0,0004	0,5749	0,5750	0,0098	0,2048	0,2050	0,5430	8,4429	8,4000	0,4580	4,7005	4,6710
		300	0,0026	0,4092	0,4090	0,0012	0,1419	0,1420	0,1070	0,6234	0,6020	0,1060	0,4777	0,4620
		500	0,0002	0,3125	0,3120	0,0042	0,1036	0,1040	0,0404	0,3383	0,3330	0,0408	0,2874	0,2840
20	45	50	0,0901	1,3330	1,2500	0,0859	0,4615	0,4540	4,4600	16,6039	15,2000	3,1600	12,9706	12,5000
		150	0,0360	0,7935	0,7730	0,0337	0,2630	0,2610	3,2200	18,8448	18,2000	2,2600	11,4691	11,2000
		300	0,0215	0,5934	0,5840	0,0215	0,1864	0,1850	0,7380	5,2919	5,1700	0,5950	2,9908	2,9100
		500	0,0080	0,4307	0,4290	0,0076	0,1358	0,1360	0,2580	1,8690	1,8300	0,2280	1,1691	1,1400
30	15	50	0,0083	1,1169	1,1200	0,0070	0,3847	0,3850	1,6900	5,8723	5,6900	1,3000	5,3230	5,1100
		150	0,0049	0,6502	0,6500	0,0114	0,2131	0,2130	0,1940	1,6261	1,6200	0,1600	1,3032	1,2900
		300	0,0024	0,4502	0,4500	0,0023	0,1562	0,1560	0,0589	0,2887	0,2840	0,0547	0,3119	0,3060
		500	0,0010	0,3535	0,3540	0,0026	0,1232	0,1230	0,0328	0,1735	0,1710	0,0299	0,2007	0,1980
45	15	50	0,1096	1,7533	1,6700	0,0930	0,5129	0,5040	4,9638	11,9741	11,2000	3,2081	10,9038	10,3000
		150	0,0364	1,0264	1,0100	0,0415	0,3029	0,3000	2,9100	11,5595	11,3000	1,7400	8,2950	8,0500
		300	0,0155	0,7531	0,7490	0,0087	0,2131	0,2130	0,5390	2,6617	2,6200	0,3360	1,8125	1,7700
		500	0,0098	0,5700	0,5680	0,0084	0,1663	0,1660	0,1560	1,0818	1,0700	0,1190	0,7836	0,7720
30	15	50	0,0271	1,2916	1,2800	0,0095	0,4589	0,4590	1,9700	3,4193	3,3300	0,9130	3,8196	3,6700
		150	0,0052	0,7215	0,7210	0,0018	0,2356	0,2360	0,1210	0,2839	0,2800	0,0905	0,4014	0,3880
		300	0,0013	0,4977	0,4970	0,0067	0,1706	0,1710	0,0673	0,1810	0,1790	0,0389	0,2555	0,2520
		500	0,0002	0,3798	0,3800	0,0009	0,1363	0,1360	0,0290	0,1359	0,1350	0,0196	0,1921	0,1910
45	15	50	0,1071	2,1348	2,0700	0,0427	0,6847	0,6830	8,4800	9,7980	9,2200	3,4109	10,6531	9,9100
		150	0,0493	1,3729	1,3500	0,0363	0,3550	0,3530	3,9100	7,6968	7,5400	1,3500	6,5276	6,3400
		300	0,0180	1,0729	1,0700	0,0120	0,2516	0,2510	1,0300	4,5218	4,5000	0,2750	2,4725	2,4500
		500	0,0054	0,7874	0,7870	0,0016	0,1959	0,1960	0,2740	0,8295	0,8230	0,1280	0,7793	0,7650

Fonte: o autor

Figura 6 – **REQM** dos parâmetros no esquema de ativação aleatória com latência log-normal e log-t-Student com três graus de liberdade sob diferentes níveis de censura e cura.

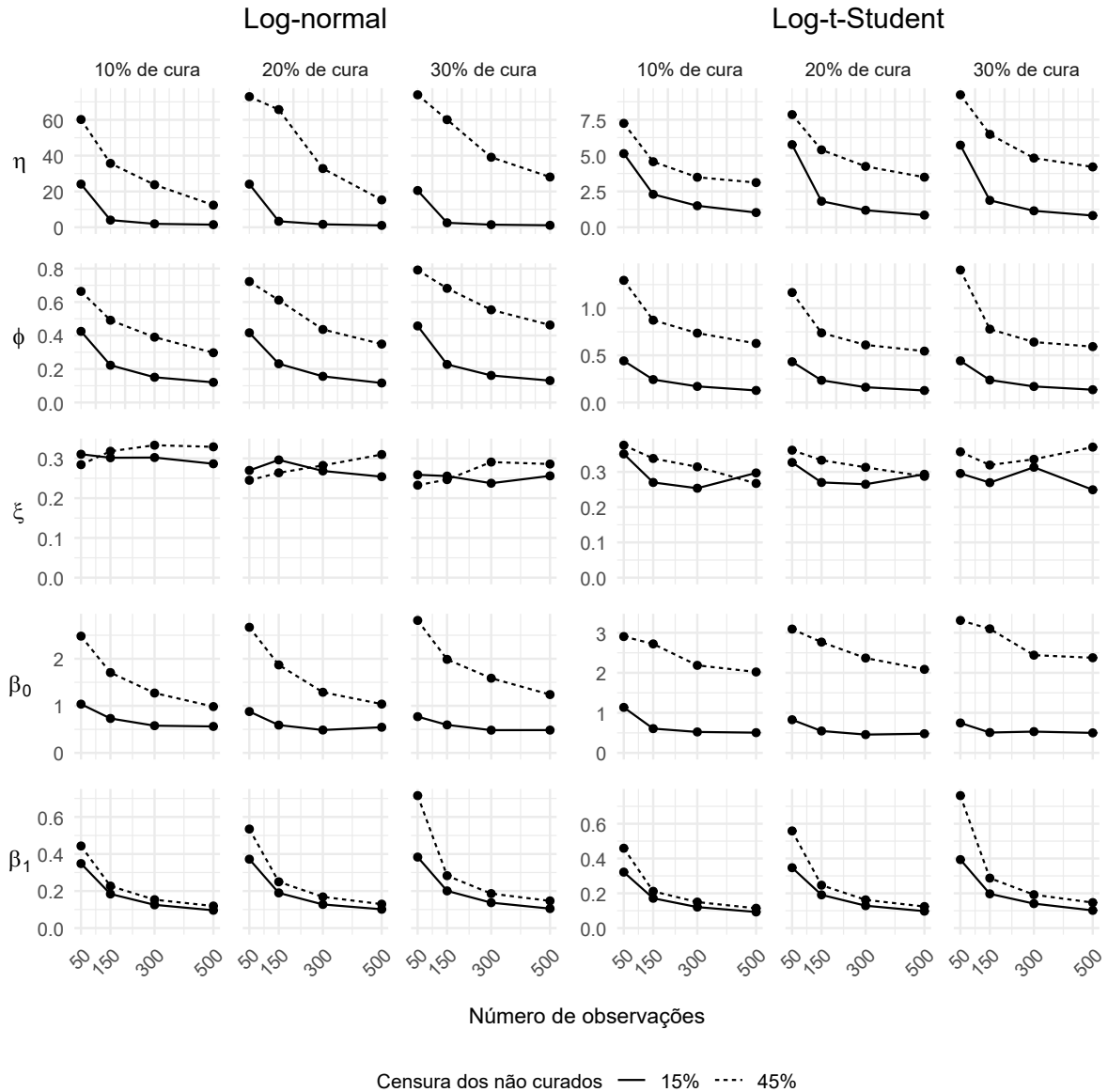


Fonte: o autor

Os resultados das simulações com ativação aleatória são apresentados nas Tabelas 12 e 13, juntamente com a Figura 6. Observa-se que o tamanho amostral impacta as métricas de **VRA**, **DP** e **REQM**, de modo que, quanto maior o tamanho da amostra, menores são os valores dessas métricas. Por outro lado, o percentual de censura afeta as métricas de forma oposta ao tamanho amostral: quanto maior o percentual de censura, maiores são os valores das métricas, embora esse efeito se reduza à medida que o tamanho amostral aumenta. Destaca-se ainda a semelhança entre os resultados dos estimadores de latência para as duas distribuições. No entanto, os estimadores de incidência, representados por β , apresentam valores elevados em

ambas as distribuições, especialmente nas métricas **DP** e **REQM**.

Figura 7 – **REQM** dos parâmetros no esquema de ativação geométrica com latência log-normal e log-*t*-Student com três graus de liberdade sob diferentes níveis de censura e cura.



Fonte: o autor

De forma semelhante aos resultados obtidos no esquema de primeira ativação, o esquema de ativação geométrica, conforme apresentado nas Tabelas 14 e 15, juntamente com a Figura 7, mostra que o estimador do parâmetro η apresenta métricas de **VRA**, **DP** e **REQM** elevadas, especialmente **DP** e **REQM**, com valores ainda maiores do que os observados no esquema de primeira ativação. Assim como nesse último, a distribuição log-*t*-Student com três graus de liberdade apresenta métricas menores em comparação à distribuição log-normal.

Observa-se, em ambas as distribuições, os mesmos comportamentos já identificados: quanto

Tabela 14 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação geométrica com latência log-normal, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			ξ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	1,2224	24,0965	23,3097	0,0577	0,4246	0,4207	0,4212	0,3105	0,2281	0,0302	1,0357	1,0315	0,0643	0,3482	0,3236
		150	0,2033	4,0176	3,8872	0,0174	0,2225	0,2218	0,2774	0,3017	0,2679	0,0028	0,7313	0,7313	0,0228	0,1841	0,1784
		300	0,0895	1,9276	1,8750	0,0095	0,1506	0,1503	0,3836	0,3022	0,2336	0,0595	0,5778	0,5476	0,0101	0,1255	0,1239
		500	0,0573	1,4703	1,4421	0,0090	0,1203	0,1200	0,3105	0,2865	0,2408	0,0420	0,5620	0,5468	0,0061	0,0961	0,0953
	45	50	5,0840	60,1662	54,5353	0,0943	0,6636	0,6569	0,3746	0,2843	0,2139	0,1599	2,4803	2,4304	0,0782	0,4432	0,4147
		150	2,2738	35,6773	33,8190	0,0710	0,4912	0,4861	0,5187	0,3180	0,1841	0,0037	1,7058	1,7058	0,0296	0,2269	0,2191
		300	1,2940	23,7585	22,8616	0,0606	0,3901	0,3854	0,5200	0,3332	0,2084	0,0314	1,2716	1,2679	0,0123	0,1533	0,1514
		500	0,6139	12,3745	11,9884	0,0390	0,2971	0,2946	0,5094	0,3291	0,2083	0,0639	0,9835	0,9634	0,0084	0,1191	0,1179
	15	50	0,9100	24,0796	23,6470	0,0223	0,4164	0,4158	0,2029	0,2698	0,2500	0,0782	0,8784	0,8628	0,0648	0,3720	0,3487
		150	0,1579	3,3482	3,2540	0,0211	0,2313	0,2304	0,3715	0,2962	0,2307	0,0805	0,5904	0,5654	0,0214	0,1899	0,1851
		300	0,0590	1,6614	1,6350	0,0014	0,1561	0,1561	0,3156	0,2685	0,2172	0,0829	0,4856	0,4531	0,0115	0,1276	0,1255
		500	0,0209	1,0478	1,0426	0,0035	0,1164	0,1163	0,1779	0,2541	0,2381	0,0223	0,5450	0,5430	0,0062	0,1015	0,1008
20	45	50	6,2876	72,9191	65,7973	0,0923	0,7224	0,7165	0,2684	0,2451	0,2051	0,3017	2,6688	2,5919	0,0995	0,5347	0,4963
		150	4,2153	65,7029	62,2338	0,1301	0,6113	0,5974	0,3592	0,2638	0,1933	0,1145	1,8691	1,8535	0,0300	0,2498	0,2425
		300	1,6100	32,7919	31,7901	0,0597	0,4360	0,4319	0,3468	0,2823	0,2228	0,0090	1,2881	1,2880	0,0147	0,1685	0,1659
		500	0,7712	15,2952	14,8020	0,0439	0,3489	0,3462	0,4756	0,3098	0,1986	0,0830	1,0370	1,0222	0,0090	0,1292	0,1279
	15	50	0,8477	20,5116	20,0699	0,0337	0,4571	0,4559	0,2512	0,2588	0,2263	0,0378	0,7698	0,7681	0,0658	0,3833	0,3600
		150	0,1086	2,5030	2,4436	0,0019	0,2273	0,2273	0,1520	0,2556	0,2440	0,0094	0,5946	0,5945	0,0223	0,2008	0,1958
		300	0,0531	1,4543	1,4299	0,0018	0,1615	0,1615	0,1940	0,2376	0,2169	0,0603	0,4827	0,4758	0,0136	0,1372	0,1344
		500	0,0423	1,1181	1,0980	0,0101	0,1307	0,1303	0,1861	0,2561	0,2386	0,0386	0,4840	0,4812	0,0059	0,1052	0,1045
30	45	50	6,3987	73,9517	66,6763	0,0220	0,7912	0,7910	0,0506	0,2327	0,2313	0,5106	2,8148	2,7279	0,1283	0,7152	0,6676
		150	4,3078	60,0691	56,0774	0,0960	0,6817	0,6749	0,1467	0,2467	0,2356	0,2681	1,9867	1,9531	0,0358	0,2830	0,2738
		300	2,5107	39,1002	37,0321	0,1190	0,5534	0,5405	0,3682	0,2908	0,2251	0,1013	1,5867	1,5807	0,0184	0,1861	0,1824
		500	1,5566	28,0611	26,9615	0,0999	0,4627	0,4518	0,2750	0,2859	0,2507	0,1103	1,2392	1,2301	0,0079	0,1468	0,1460

Fonte: o autor

Tabela 15 – **VRA**, **REQM** e **DP** das estimativas dos parâmetros no esquema de ativação geométrica com latência log-t-Student com três graus de liberdade, sob diferentes níveis de censura e cura.

fc(%)	pcNC(%)	n	η			ϕ			ξ			β_0			β_1		
			VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP	VRA	REQM	DP
10	15	50	0.2076	5.1361	5.0304	0.1367	0.4414	0.4197	0.5868	0.3509	0.1925	0.0466	1.1362	1.1272	0.0501	0.3218	0.3058
		150	0.0568	2.2970	2.2795	0.0502	0.2428	0.2376	0.2401	0.2701	0.2420	0.0203	0.6060	0.6028	0.0186	0.1714	0.1673
		300	0.0456	1.4996	1.4823	0.0329	0.1707	0.1675	0.2346	0.2537	0.2249	0.0275	0.5228	0.5160	0.0085	0.1207	0.1195
		500	0.0160	1.0273	1.0243	0.0197	0.1277	0.1262	0.4718	0.2971	0.1807	0.1072	0.5053	0.3844	0.0064	0.0926	0.0917
45	50	0.0178	7.2445	7.2444	0.2424	1.2955	1.2727	0.5423	0.3757	0.2600	0.0550	2.9064	2.9017	0.0841	0.4591	0.4272	
	150	0.0725	4.5704	4.5562	0.0241	0.8727	0.8724	0.4404	0.3381	0.2566	0.2126	2.7220	2.6432	0.0273	0.2114	0.2042	
	300	0.0712	3.4868	3.4688	0.0137	0.7356	0.7355	0.4147	0.3143	0.2362	0.1329	2.1868	2.1487	0.0137	0.1495	0.1469	
	500	0.0717	3.1165	3.0960	0.0121	0.6265	0.6264	0.2528	0.2673	0.2356	0.1298	2.0211	1.9818	0.0097	0.1139	0.1122	
20	15	50	0.1757	5.7592	5.6920	0.1038	0.4319	0.4192	0.5180	0.3269	0.1995	0.1160	0.8276	0.7913	0.0545	0.3471	0.3296
		150	0.0449	1.8161	1.8022	0.0395	0.2353	0.2319	0.3418	0.2701	0.2091	0.0994	0.5470	0.5061	0.0192	0.1911	0.1872
		300	0.0162	1.1873	1.1846	0.0206	0.1618	0.1605	0.3275	0.2650	0.2084	0.1032	0.4570	0.4029	0.0089	0.1287	0.1275
		500	0.0153	0.8504	0.8470	0.0143	0.1271	0.1263	0.4114	0.2933	0.2091	0.1309	0.4780	0.3920	0.0066	0.0972	0.0963
45	50	0.0553	7.8534	7.8490	0.0801	1.1666	1.1639	0.4211	0.3616	0.2940	0.2397	3.0926	3.0519	0.0990	0.5587	0.5225	
	150	0.1488	5.3966	5.3453	0.0580	0.7382	0.7359	0.3655	0.3332	0.2786	0.3453	2.7674	2.6718	0.0309	0.2466	0.2387	
	300	0.1453	4.2431	4.1806	0.0825	0.6104	0.6048	0.2824	0.3130	0.2793	0.3232	2.3689	2.2707	0.0136	0.1623	0.1600	
	500	0.1304	3.4912	3.4299	0.0668	0.5448	0.5407	0.0086	0.2874	0.2874	0.3777	2.0868	1.9318	0.0084	0.1238	0.1227	
30	15	50	0.2019	5.7252	5.6359	0.1041	0.4415	0.4291	0.4388	0.2956	0.1982	0.1561	0.7474	0.7152	0.0666	0.3937	0.3705
		150	0.0483	1.8770	1.8615	0.0439	0.2382	0.2341	0.3457	0.2696	0.2069	0.1566	0.5084	0.4595	0.0250	0.1970	0.1905
		300	0.0199	1.1482	1.1439	0.0137	0.1700	0.1695	0.5022	0.3132	0.1872	0.2644	0.5315	0.3839	0.0108	0.1407	0.1390
		500	0.0076	0.8175	0.8167	0.0059	0.1353	0.1352	0.1849	0.2493	0.2315	0.0594	0.4993	0.4925	0.0047	0.1017	0.1013
45	50	0.1558	9.2332	9.2007	0.1075	1.4050	1.4010	0.2303	0.3569	0.3378	0.5301	3.3090	3.2261	0.1397	0.7618	0.7087	
	150	0.2108	6.4768	6.3907	0.0769	0.7780	0.7742	0.1193	0.3195	0.3139	0.8611	3.0985	2.8581	0.0399	0.2874	0.2761	
	300	0.1234	4.8233	4.7839	0.0213	0.6406	0.6403	0.2459	0.3358	0.3125	0.4832	2.4409	2.3468	0.0154	0.1920	0.1896	
	500	0.1725	4.2066	4.1174	0.0873	0.5925	0.5860	0.0989	0.3703	0.3670	0.7657	2.3761	2.1245	0.0092	0.1465	0.1453	

Fonte: o autor

maior o percentual de censura, maiores são os valores das métricas; o efeito do aumento da $pcNC$ é intensificado à medida que a fc cresce; e, quanto maior o tamanho amostral, menores tendem a ser os valores das métricas, com exceção do estimador $\hat{\xi}$, que não apresenta uma tendência clara de redução conforme o tamanho amostral aumenta.

4.1.4 Conclusão

De modo geral, ao analisar as diversas simulações sob diferentes Cenários de fc , $pcNC$ e tamanho amostral, observa-se que o aumento do tamanho amostral tende a produzir estimativas menos viesadas e com menor variabilidade, com exceção do estimador para o parâmetro ξ no esquema de ativação geométrica. Nota-se também que, em geral, o aumento da fc intensifica o crescimento das métricas decorrente do aumento da $pcNC$.

O estimador da mediana, η , apresentou métricas elevadas, em especial alta variabilidade DP e $REQM$ em todos os esquemas, Cenários e distribuições de latência, com desempenho melhor sob a distribuição log-t-Student com três graus de liberdade em comparação à log-normal. A única exceção foi observada no esquema de ativação aleatória, onde os resultados de $\hat{\eta}$ foram próximos dos demais parâmetros nas duas distribuições e 2 Cenários.

Quanto às covariáveis, os resultados foram semelhantes entre os diferentes Cenários, exceto no esquema de ativação aleatória com covariável contínua, no qual os estimadores dos parâmetros β apresentaram métricas superiores às dos demais estimadores, incluindo os β sob a covariável de grupo no mesmo esquema.

4.2 RESIDUOS DE COX-SNELL

Após o ajuste do modelo estatístico proposto, torna-se necessário avaliar a qualidade do ajuste a fim de identificar possíveis fugas das suposições feitas para o modelo. Nesse sentido, diversas técnicas foram desenvolvidas para avaliar a qualidade do ajuste. Com base nos resíduos é possível construir gráficos que comparam os resíduos observados com os esperados e permitem identificar a existência de observações atípicas e violações das suposições do modelo, como inadequação da distribuição assumida para a variável resposta, escolha incorreta da função de ligação ou omissão de covariáveis relevantes. No contexto da análise de sobrevivência, a análise de resíduos é bastante abrangente, com diferentes tipos de resíduos disponíveis, entre eles, Cox-Snell (COX; SNELL, 1968), desvio (MCCULLAGH; NELDER, 1983), martin-

gale (THERNEAU; GRAMBSCH; FLEMING, 1990) score (THERNEAU, 2000) e Schoenfeld (WILEYTO et al., 2013).

A literatura sobre diagnóstico em modelos de fração de cura é relativamente recente e ainda pouco explorada. Especificamente quanto à análise de resíduos, alguns exemplos incluem, Peng e Taylor (2017), que utiliza resíduos martingale, de desvio, Cox–Snell e suas versões modificadas em um modelo com tempo de vida não paramétrico, sob a suposição de riscos proporcionais de Cox, suposição verificada por meio da análise dos resíduos; Scolas et al. (2018), que avalia o desempenho dos resíduos de Cox–Snell (global e local) e do resíduo de desvio no modelo de mistura, assumindo tempo de vida com distribuição gama generalizada estendida e censura intervalar; e Fidelis, Ortega e Cordeiro (2024), que analisa o desempenho do resíduo de Cox–Snell global no modelo tempo de promoção, considerando o tempo de vida com distribuição Weibull-exponencializada.

Dessa forma, foi realizado um estudo de simulação de Monte Carlo, com 1000 réplicas, com o objetivo de avaliar o comportamento do resíduo de Cox–Snell global na detecção de especificações incorretas do modelo. As configurações utilizadas foram as mesmas da seção anterior, considerando os diferentes níveis de fc e $pcNC$, porém, com os dois tipos de covariáveis incluídos simultaneamente, ou seja, assume-se que $\theta_i = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$, com uma covariável de grupo e uma contínua, sendo os valores de $\boldsymbol{\beta}$ escolhidos de forma a produzir a fc desejada, conforme apresentado na Tabela 16. Além disso, foram considerados três tamanhos amostrais: $n = 150, 300$ e 500 .

Tabela 16 – Vetores fixados para os $\boldsymbol{\beta}$ nas simulações dos esquemas de primeira ativação, ativação aleatória e ativação geométrica

$fc(\%)$	Primeira e Aleatória	Geométrica
10	$\boldsymbol{\beta} = (1,77, 0,15, 1,5)^\top$	$\boldsymbol{\beta} = (1,88, 2, 1,5)^\top$
20	$\boldsymbol{\beta} = (0,99, 0,15, 1,5)^\top$	$\boldsymbol{\beta} = (0,98, 2, 1,5)^\top$
30	$\boldsymbol{\beta} = (0,42, 0,15, 1,5)^\top$	$\boldsymbol{\beta} = (0,33, 2, 1,5)^\top$

Fonte: o autor

4.2.1 Resíduo de Cox-Snell

O resíduo de Cox–Snell é utilizado para detectar especificações incorretas no modelo. Esse resíduo tem a forma $-\log\{S_T(t)\}$ e pela transformada integral de probabilidade segue uma distribuição exponencial com parâmetro 1. Com base nesse fato, é possível identificar

a existência de especificação inadequada quando o resíduo calculado a partir do modelo não apresenta o comportamento esperado. Uma forma gráfica de verificar isso é comparar os resíduos estimados pelo modelo no eixo x com os resíduos esperados no eixo y . Como há observações censuradas, os resíduos esperados são estimados por meio do **EKM**. Assim, o gráfico deve apresentar pontos distribuídos ao longo de uma linha com inclinação 1 e intercepto 0, sendo que desvios dessa linha indicam especificação incorreta do modelo.

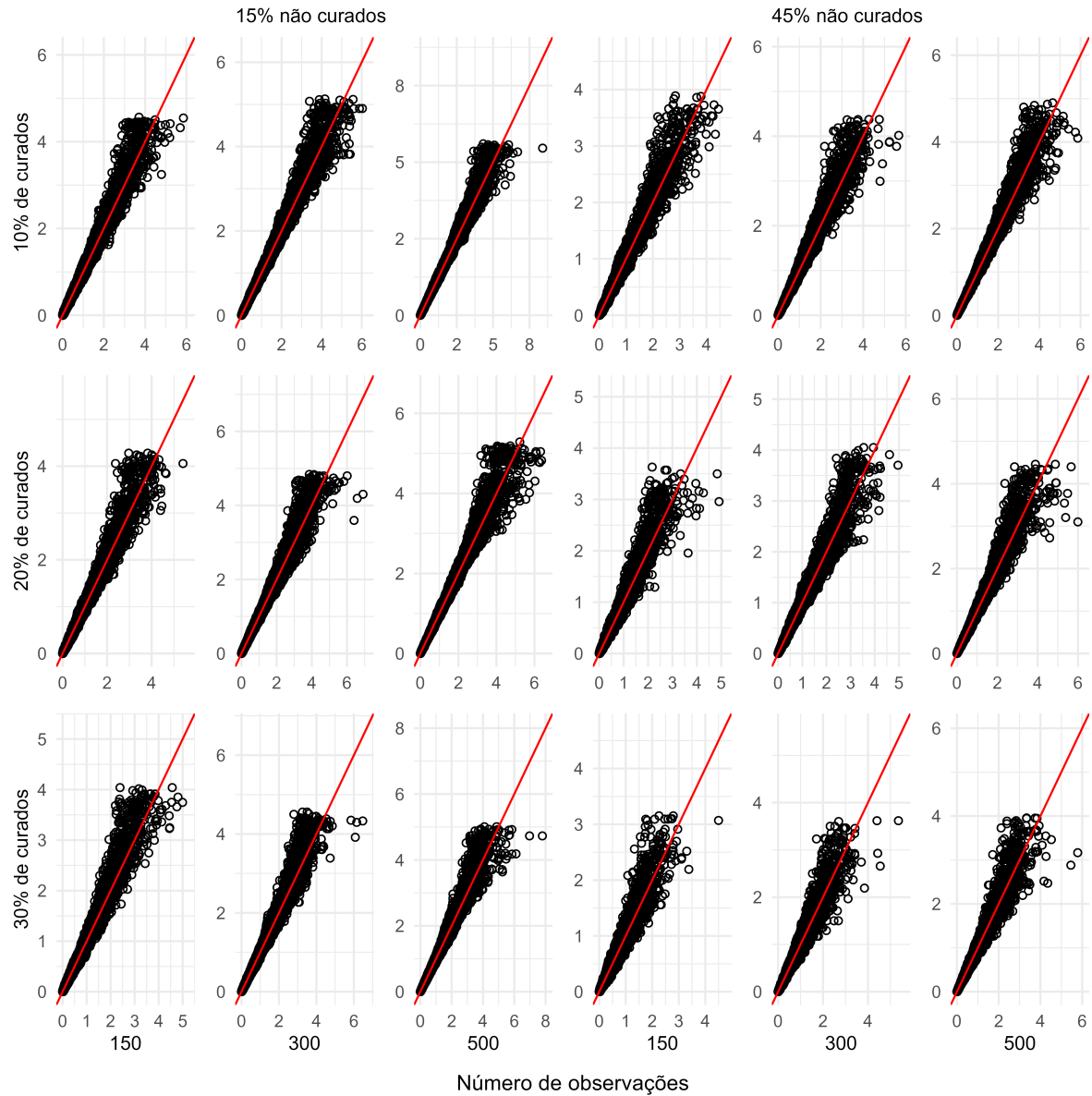
Para o modelo de fração de cura log-simétrico com diferentes esquemas de ativação, os resíduos de Cox–Snell são dados por

$$r_{cs} = -\log \left\{ P(R > M) + E_{R,M} \left[IB(S_{\tilde{Y}_r}; M - R + 1, R) I(M \geq R \geq 1) \right] \right\}.$$

Diferentes Cenários foram simulados com foco na detecção de possíveis especificações incorretas das distribuições que compõem o modelo, sendo elas: as distribuições do esquema, da incidência e da latência. O efeito da mesma especificação errada foi analisada sob diferentes distribuições do esquema de ativação e distribuições de latência. As especificações dos Cenários foram, Cenário com especificação correta (A); Cenário com especificação incorreta na distribuição da latência, em que o modelo é estimado usando a distribuição log-t-Student com três graus de liberdade para dados gerados pela distribuição log-normal (B_{lat}); Cenário com especificação incorreta na distribuição da incidência, em que o modelo é estimado usando distribuição Bernoulli com probabilidade $\theta_i = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} / (1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\})$ para dados gerados pela distribuição Poisson com média $\theta_i = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$ (B_{inc}); e Cenário com especificação incorreta no esquema de ativação, em que o modelo é estimado usando um esquema enquanto os dados foram gerados por outro (B_{esq}).

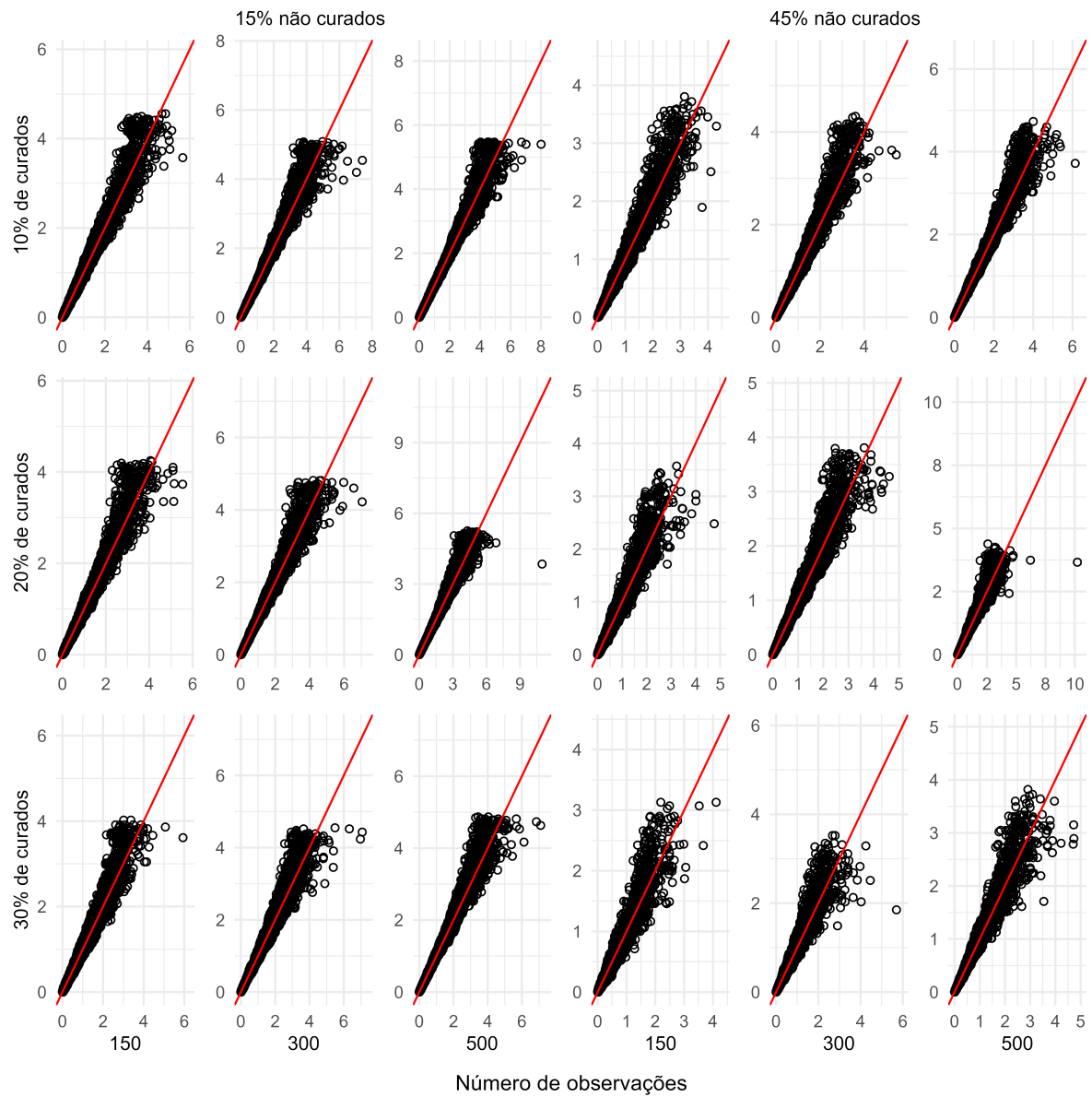
4.2.2 Resíduo de Cox-Snell com especificação correta

Figura 8 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-normal e esquema de primeira ativação.



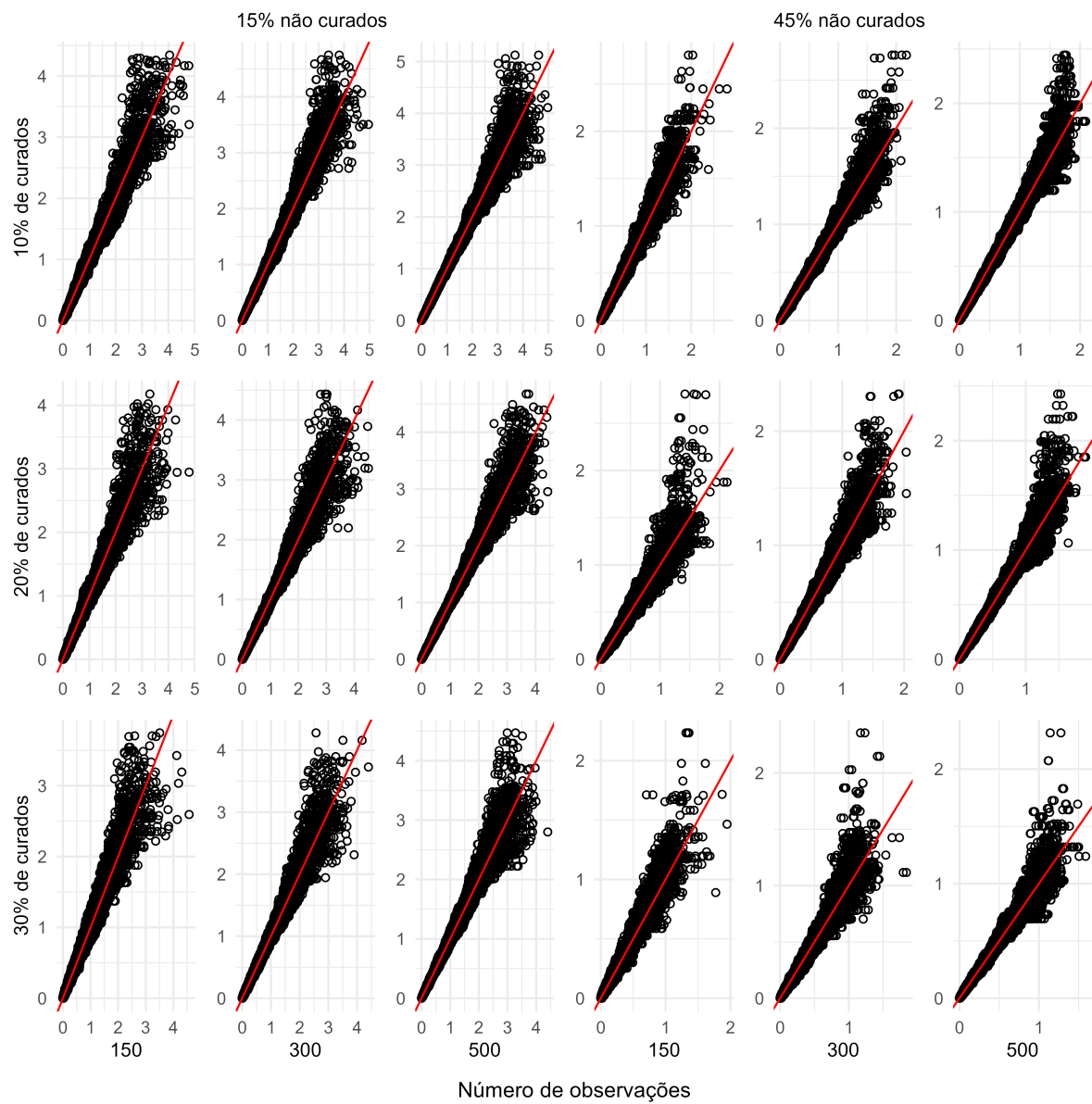
Fonte: o autor

Figura 9 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-t-Student com três graus de liberdade e esquema de primeira ativação.



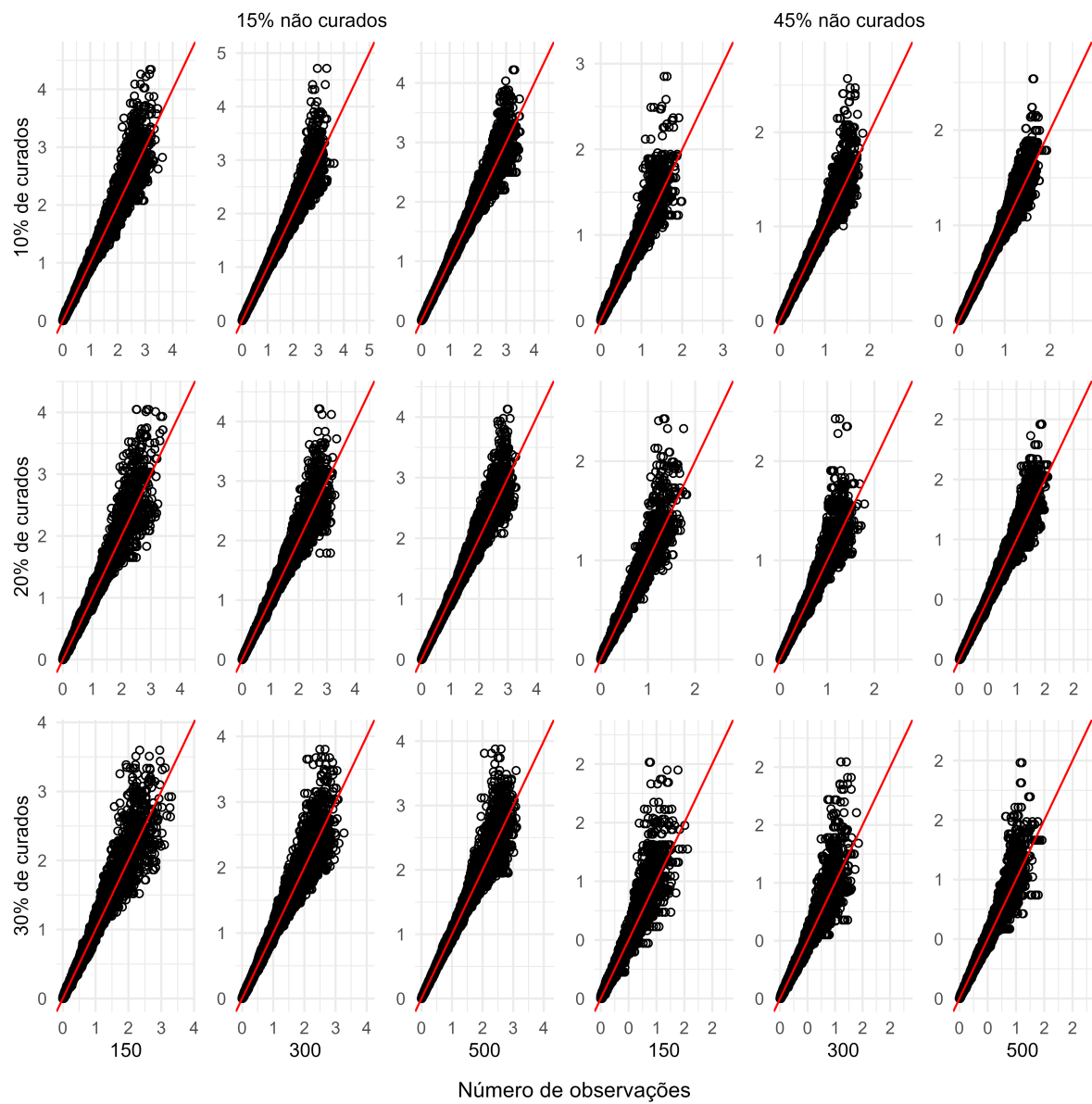
Fonte: o autor

Figura 10 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-normal e esquema de ativação aleatória.



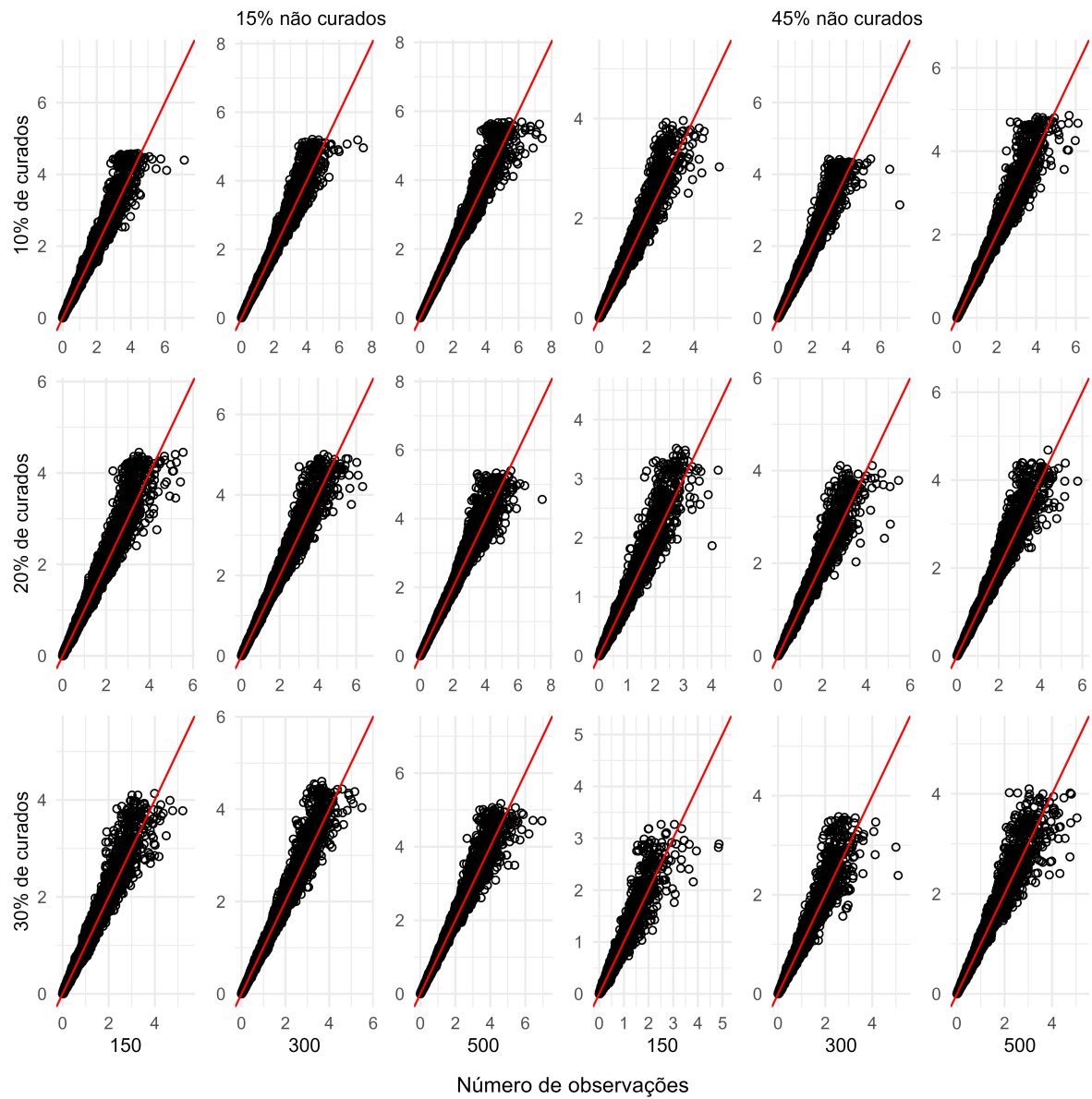
Fonte: o autor

Figura 11 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-t-Student com três graus de liberdade e esquema de ativação aleatória.



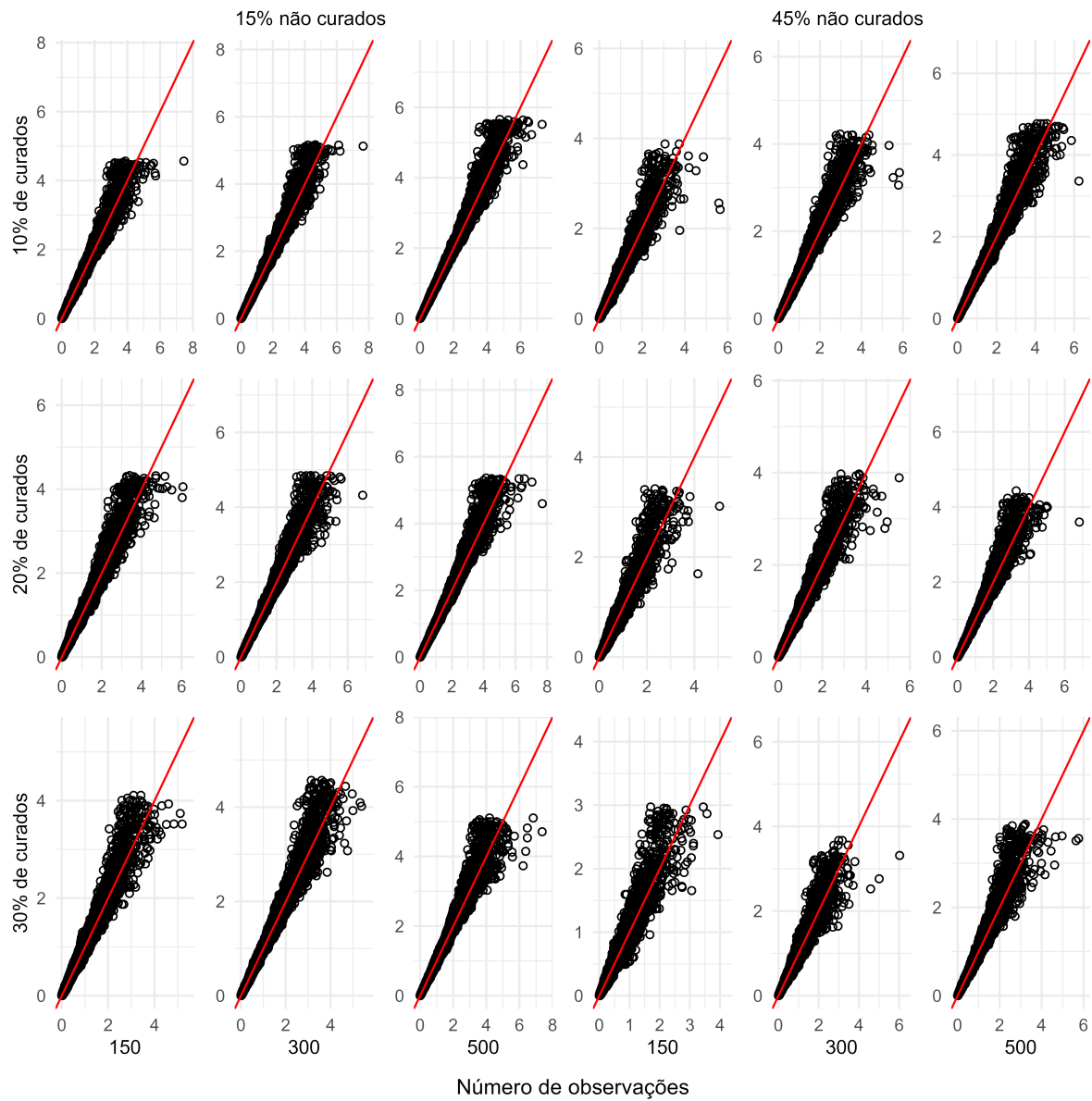
Fonte: o autor

Figura 12 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-normal e esquema de ativação geométrica.



Fonte: o autor

Figura 13 – Resíduo observado *versus* resíduo esperado para a especificação correta com distribuição log-t-Student com três graus de liberdade e esquema de ativação geométrica.



Fonte: o autor

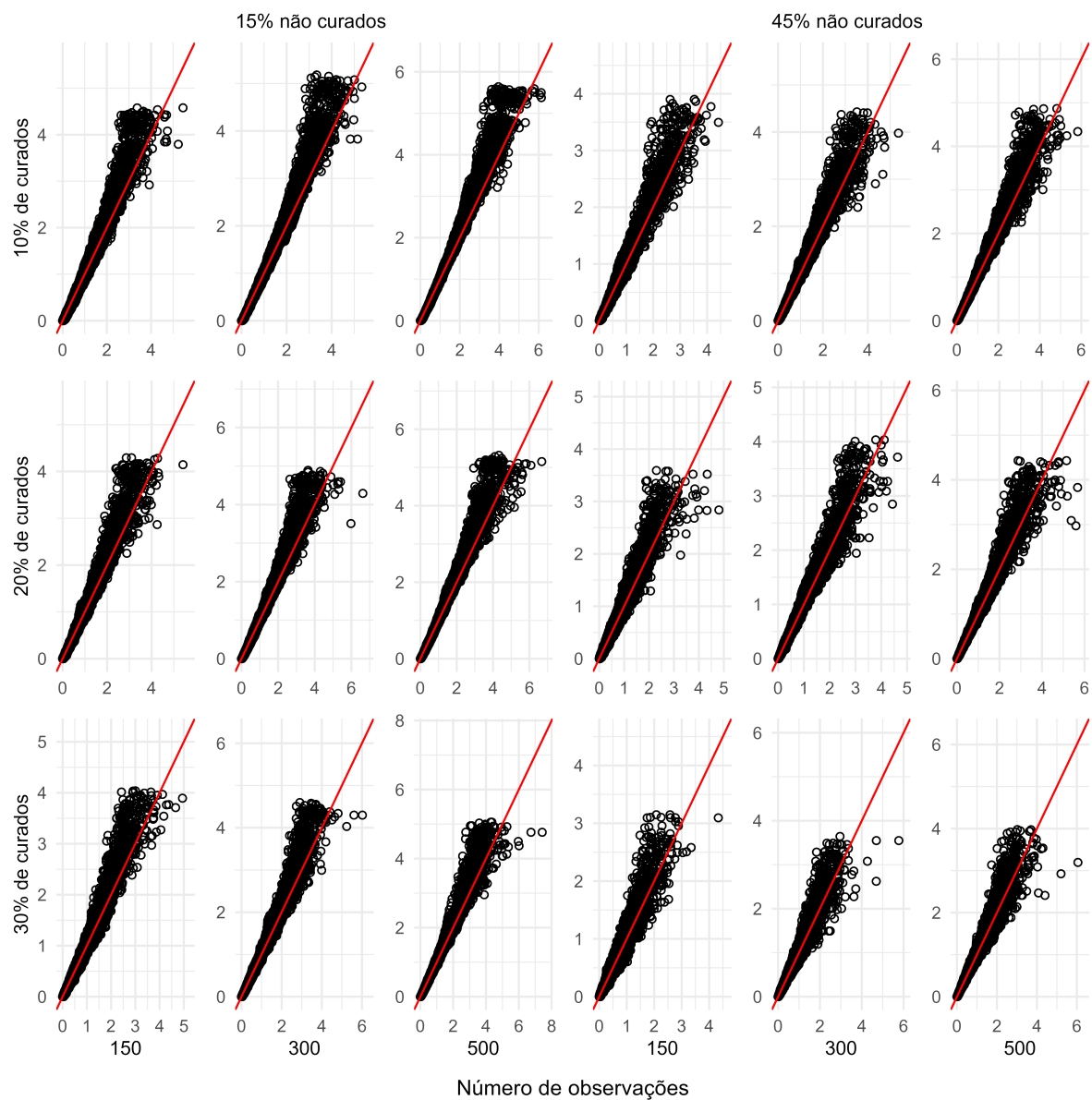
As Figuras 8 a 13 exibem os gráficos dos resíduos de Cox-Snell para as distribuições de latência log-normal e log-t-Student com três graus de liberdade, no Cenário A, considerando os diferentes esquemas de ativação. Observa-se que, para ambas as distribuições, os pontos estão bem distribuídos ao redor da linha de referência, com inclinação igual a 1 e intercepto 0, exceto no esquema de ativação aleatória representado pelas Figuras 10 e 11, no qual, na região final, observa-se uma assimetria à esquerda que é agravada nos casos com maior percentual de censura.

Além disso, nota-se uma maior dispersão à medida que os valores dos resíduos aumentam,

sendo essa dispersão menor na distribuição log-t-Student com três graus de liberdade em comparação com a log-normal. Observa-se também que, com o crescimento da censura, surgem mais pontos afastados dos demais, embora proporcionalmente ainda sejam poucos. Por outro lado, à medida que o tamanho amostral aumenta, os pontos tendem a ser mais concentrados.

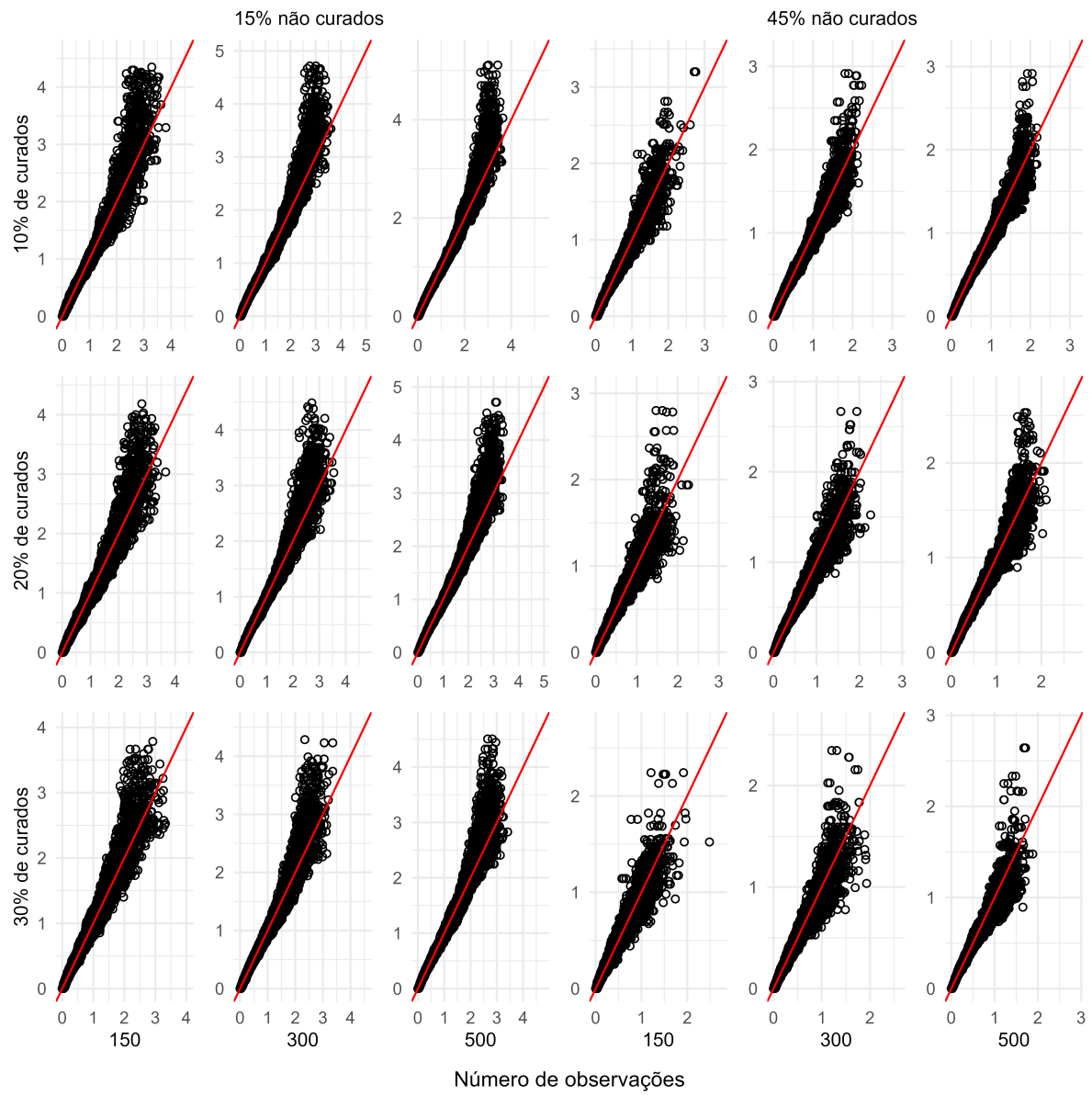
4.2.3 Resíduo de Cox-Snell com especificação incorreta na latência

Figura 14 – Resíduo observado *versus* resíduo esperado para a especificação incorreta na latência e esquema de primeira ativação.



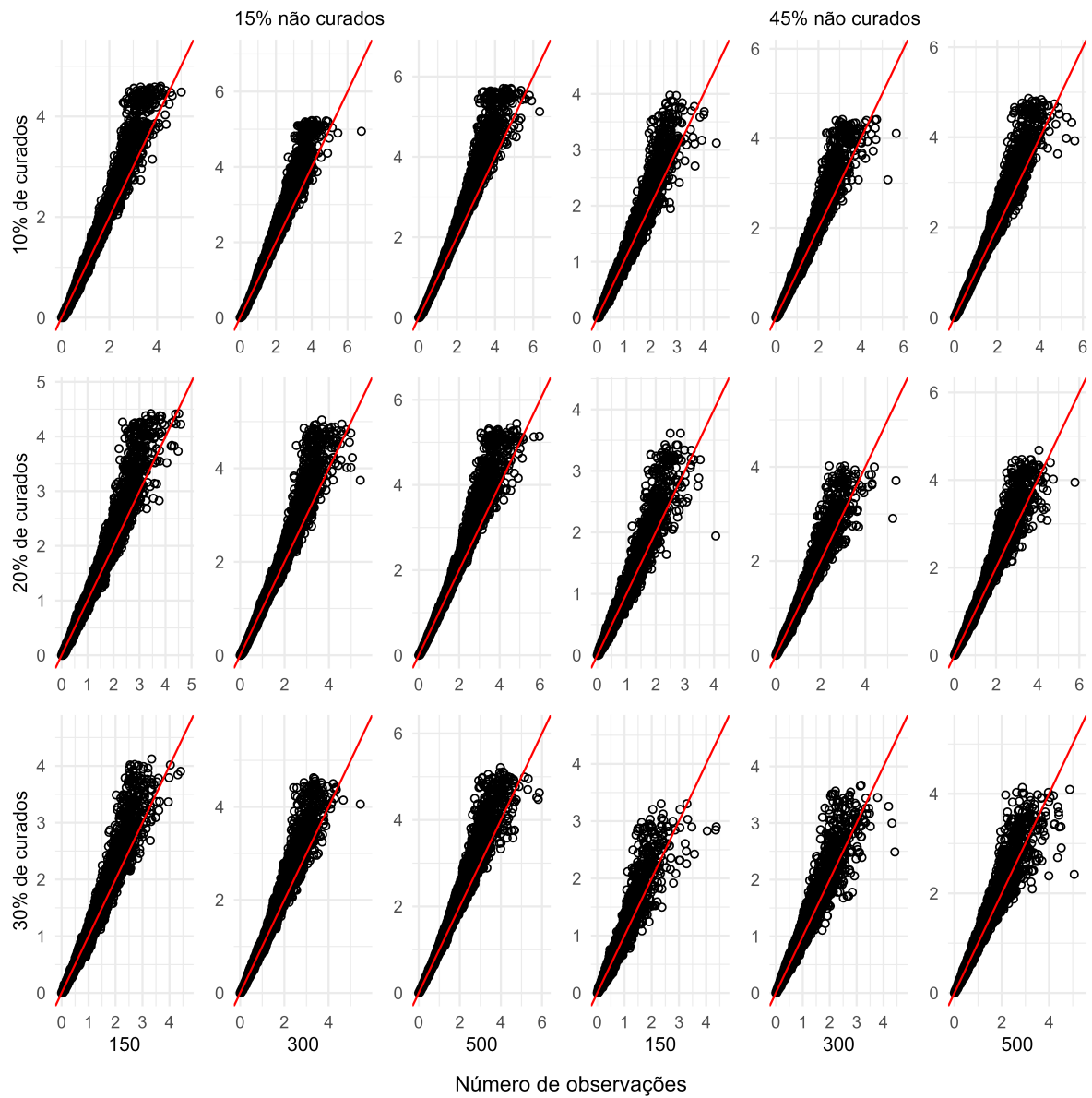
Fonte: o autor

Figura 15 – Resíduo observado *versus* resíduo esperado para a especificação errada na latência e esquema de ativação aleatória.



Fonte: o autor

Figura 16 – Resíduo observado *versus* resíduo esperado para a especificação errada na latência e esquema de ativação geométrica.



Fonte: o autor

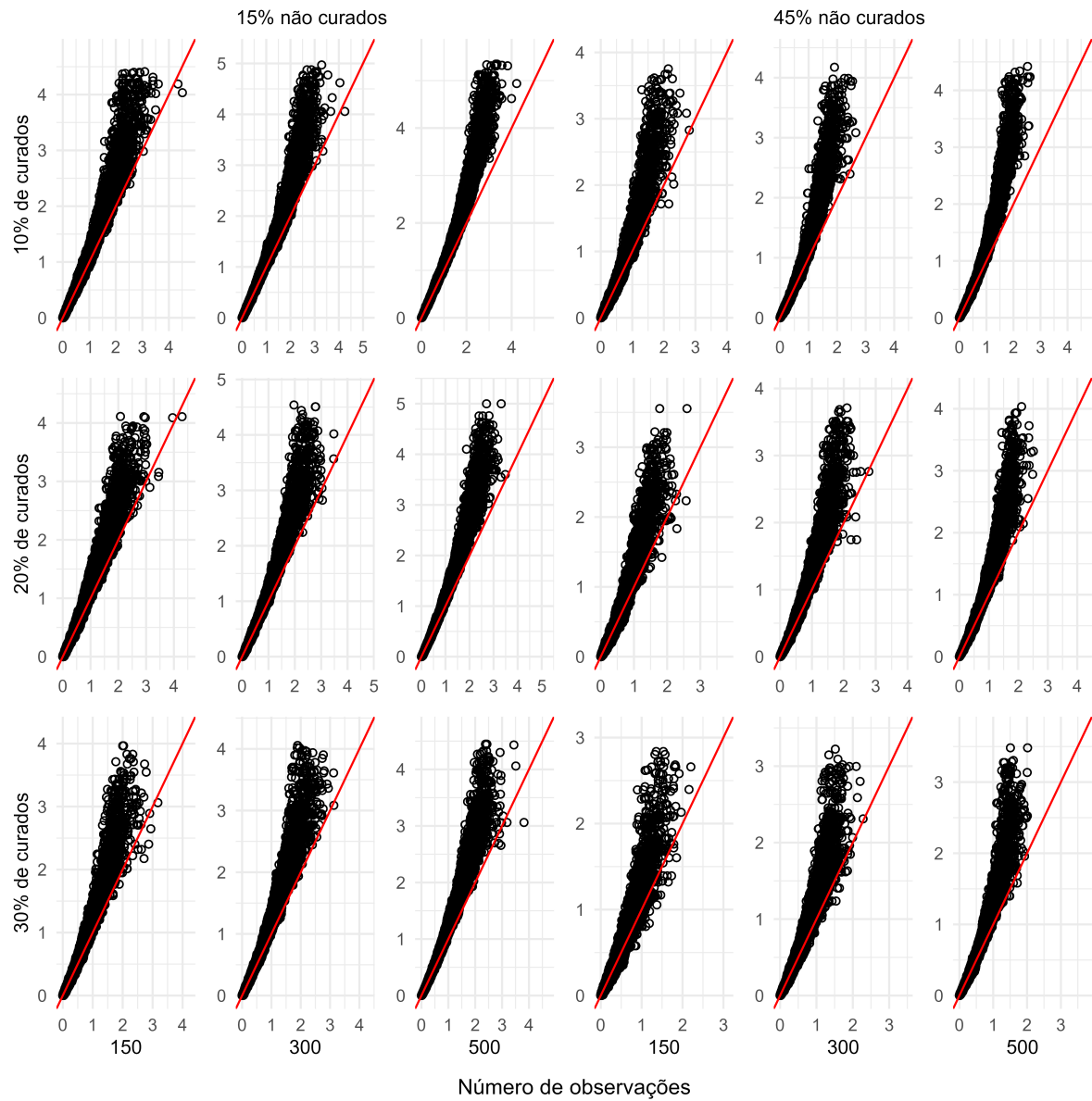
O Cenário B_{lat} nos diferentes esquemas de ativação, é representado pelas Figuras 14 a 16. Observa-se uma leve tendência dos pontos se concentrarem à esquerda da linha de referência com inclinação igual a 1 e intercepto 0, nos esquemas de primeira ativação e ativação geométrica. Além disso, no esquema de ativação aleatória, o comportamento assimétrico na região final dos pontos, torna-se mais acentuado.

Esse padrão de assimetria nos diferentes esquemas torna-se mais evidente com o aumento do tamanho amostral, embora sua intensidade diminua à medida que a censura cresce, o que pode dificultar a identificação de especificações incorretas em Cenários com alto percentual

de censura.

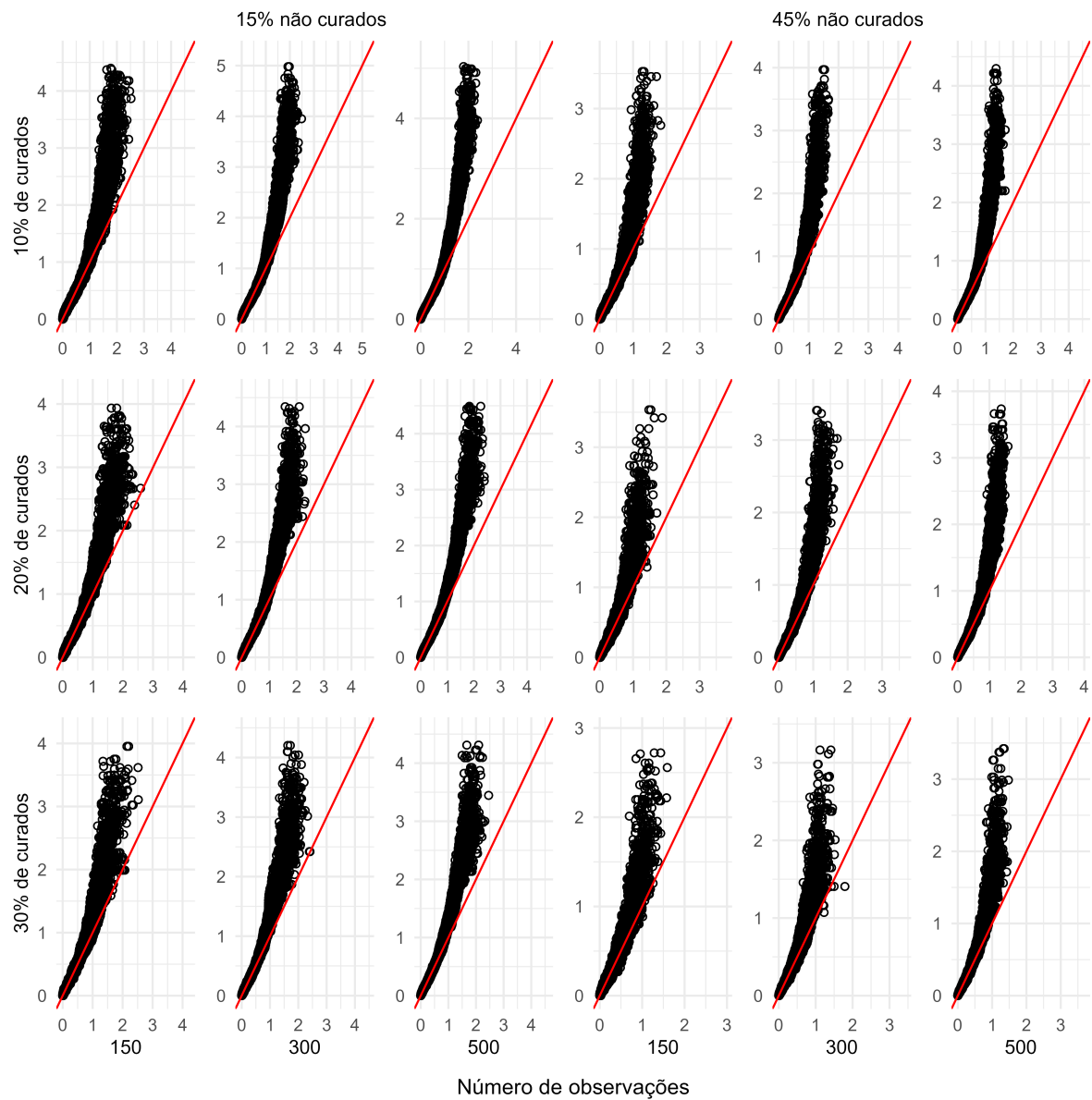
4.2.4 Resíduo de Cox-Snell com especificação incorreta na incidência

Figura 17 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-normal e esquema de primeira ativação.



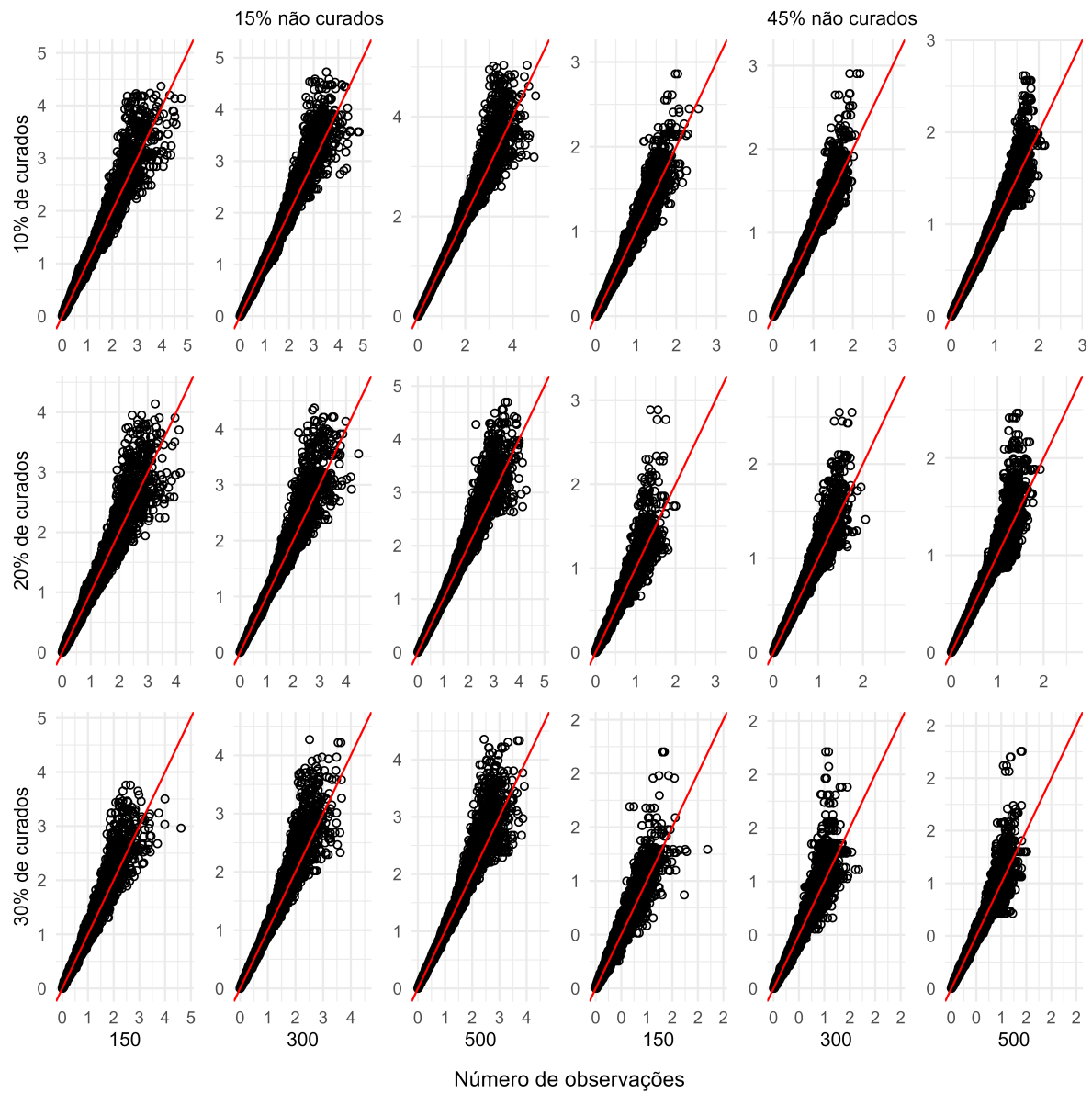
Fonte: o autor

Figura 18 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-t-Student com três graus de liberdade e esquema de primeira ativação.



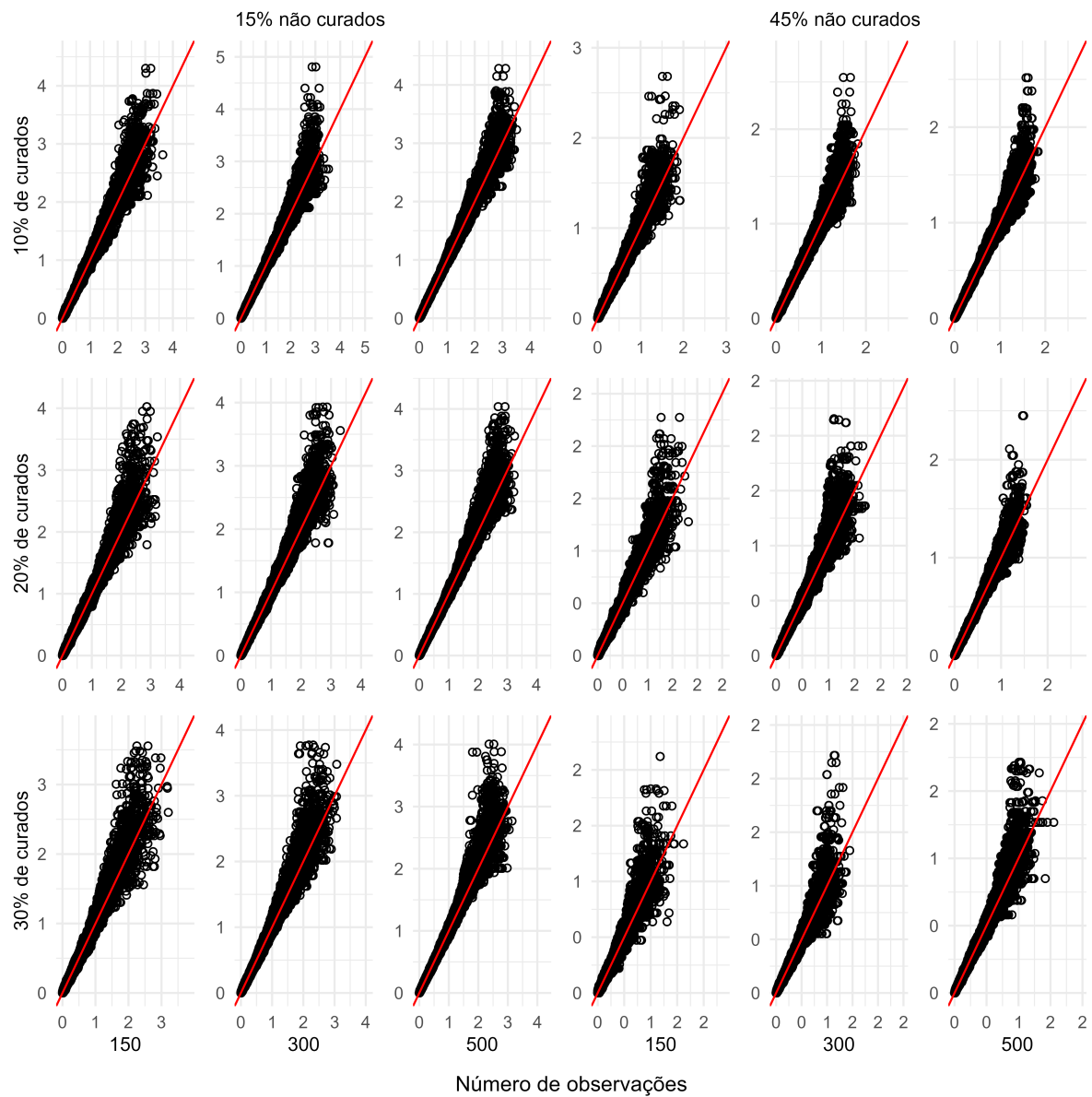
Fonte: o autor

Figura 19 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-normal e esquema de ativação aleatória.



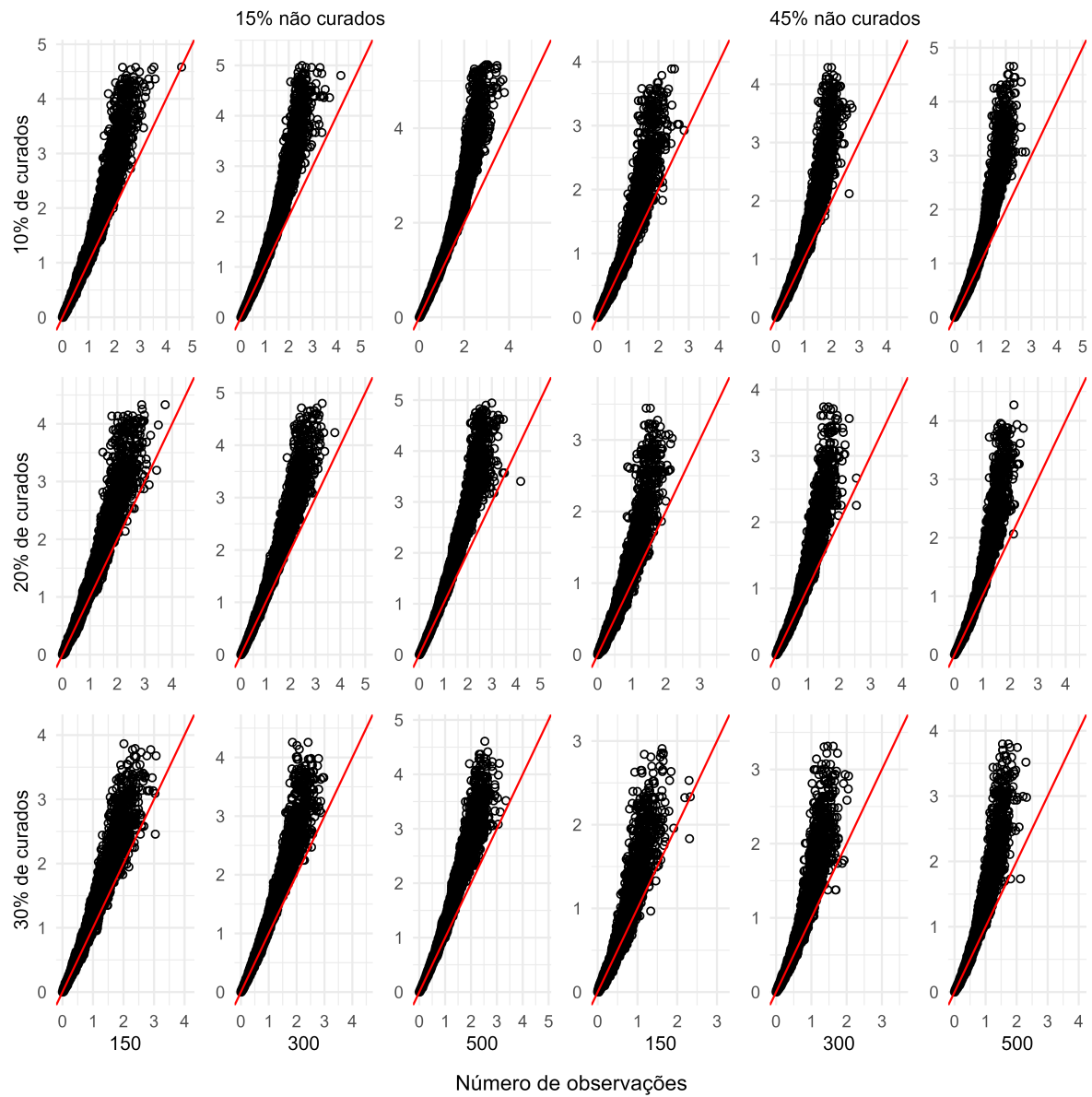
Fonte: o autor

Figura 20 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-t-Student com três graus de liberdade e esquema de ativação aleatória.



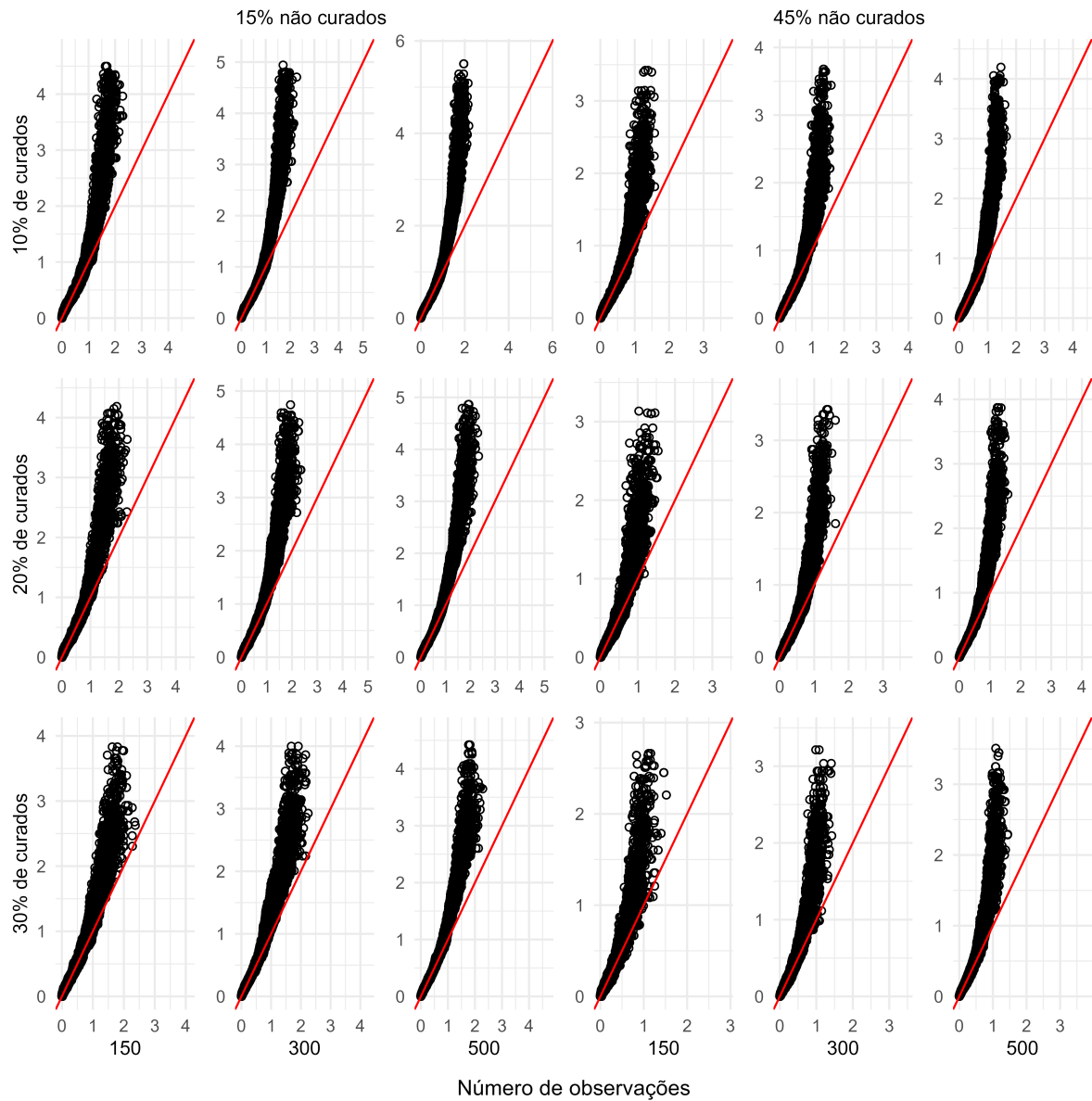
Fonte: o autor

Figura 21 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-normal e esquema de ativação geométrica.



Fonte: o autor

Figura 22 – Resíduo observado *versus* resíduo esperado para a especificação errada na incidência com distribuição log-t-Student com três graus de liberdade e esquema de ativação geométrica.



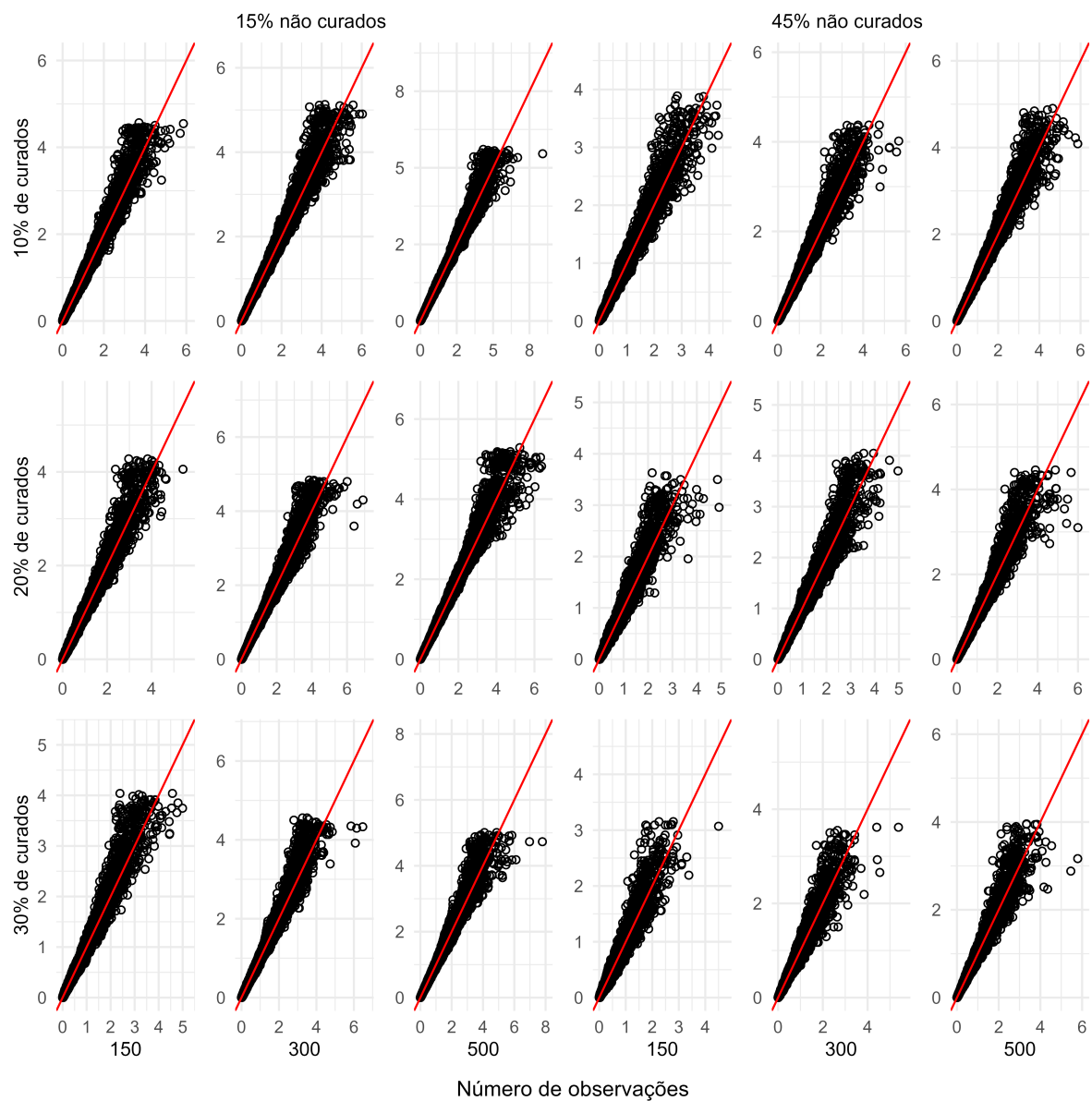
Fonte: o autor

Para o Cenário B_{inc} , nota-se a partir das Figuras 17 a 21, uma tendência de inclinação diferente de 1 em todos os esquemas de ativação, a qual se intensifica à medida que os resíduos aumentam, sendo mais pronunciada na distribuição log-t-Student com três graus de liberdade. Esse comportamento torna-se mais evidente com o aumento do tamanho amostral e do percentual de censura dos não curados (pcNC) e é parcialmente atenuado com o aumento da fração de cura (fc), embora a especificação incorreta ainda possa ser detectada. Além disso, observa-se que, no esquema de ativação aleatória, há um número maior de pontos desviando-

se da linha de referência em comparação ao Cenário com especificação correta, especialmente quando a **fc** é elevada, embora esse efeito seja suavizado com o aumento da **pcNC**.

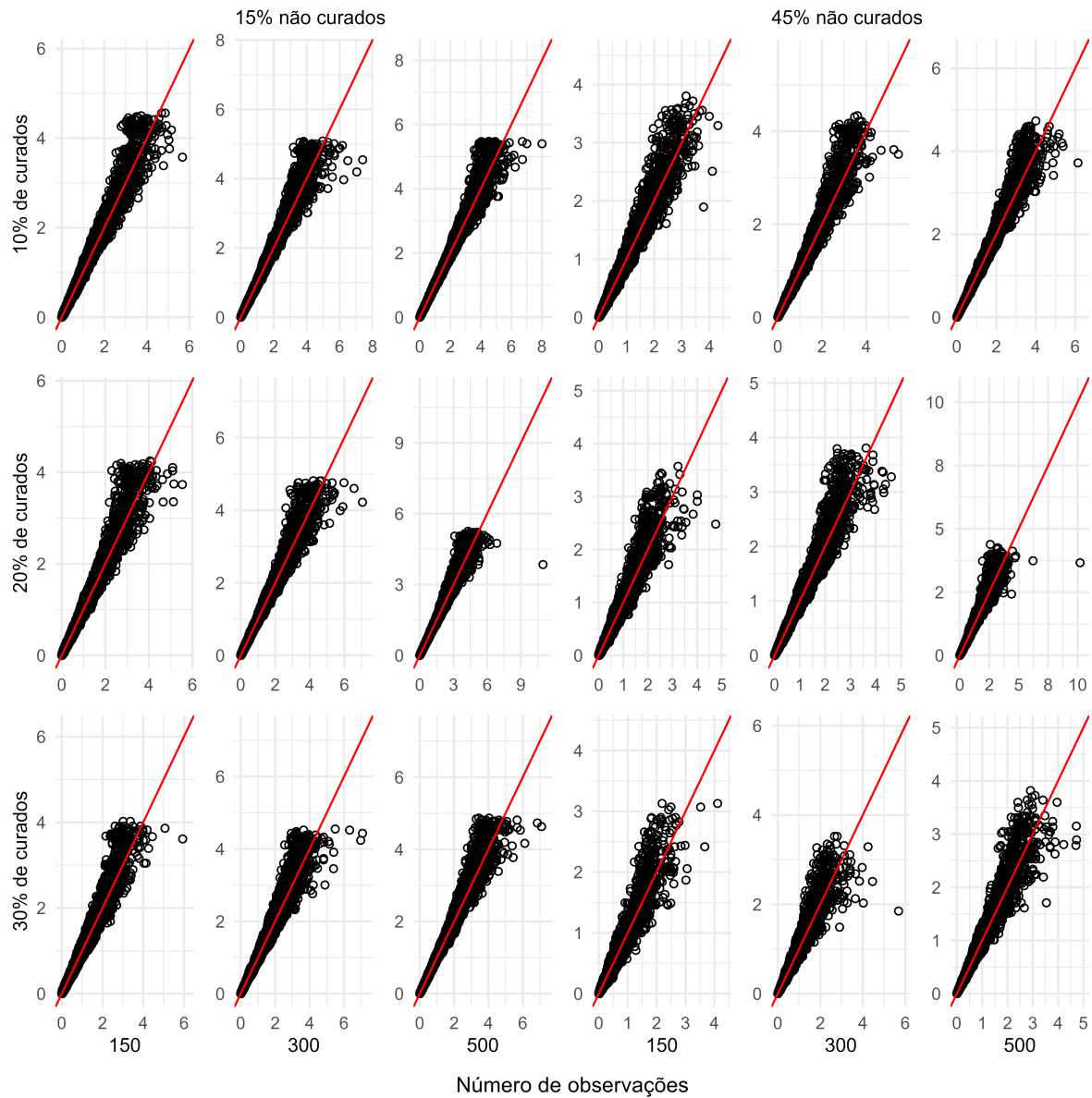
4.2.5 Resíduo de Cox-Snell com especificação incorreta no esquema

Figura 23 – Resíduo observado *versus* resíduo esperado para a especificação errada de esquema no esquema de primeira ativação com distribuição log-normal.



Fonte: o autor

Figura 24 – Resíduo observado *versus* resíduo esperado para a especificação errada de esquema no esquema de primeira ativação com distribuição log-t-Student com três graus de liberdade.

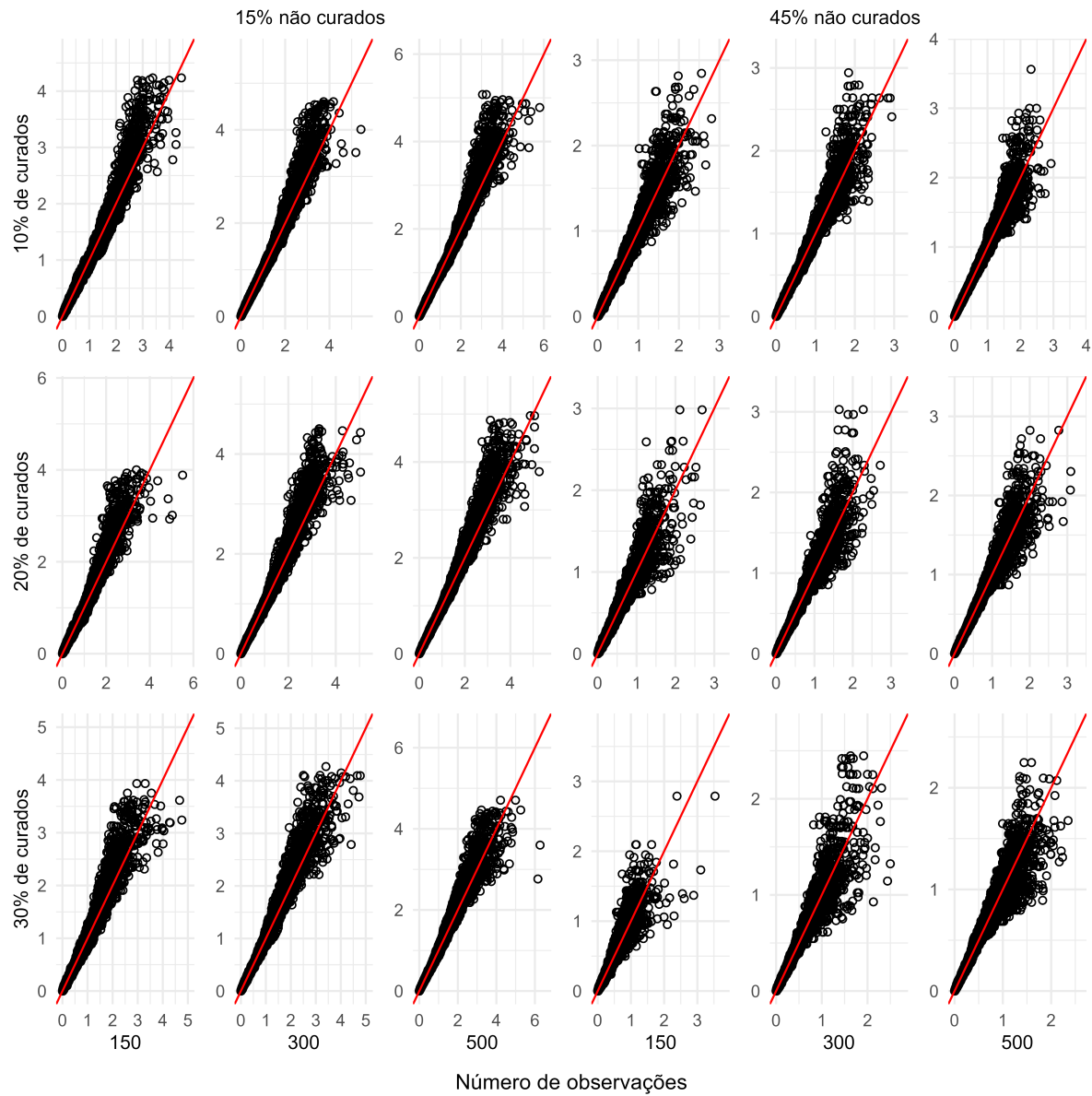


Fonte: o autor

As Figuras de 23 a 28 apresentam os resultados do Cenário B_{esq} . Das Figuras 23 e 16 nas quais os dados foram gerados com o esquema de primeira ativação, mas o modelo foi ajustado considerando o esquema de ativação geométrica (que se reduz ao de primeira ativação quando $\xi = 1$). Observa-se que, nesse caso, os resíduos não são capazes de detectar a especificação incorreta mesmo com a média das estimativas do parâmetro ξ girando em torno de 0,49 a 0,75, mostrando resultados muito próximos do caso com especificação correta, mesmo em situações com grandes tamanhos amostrais e baixos níveis de censura. Embora não mostrado

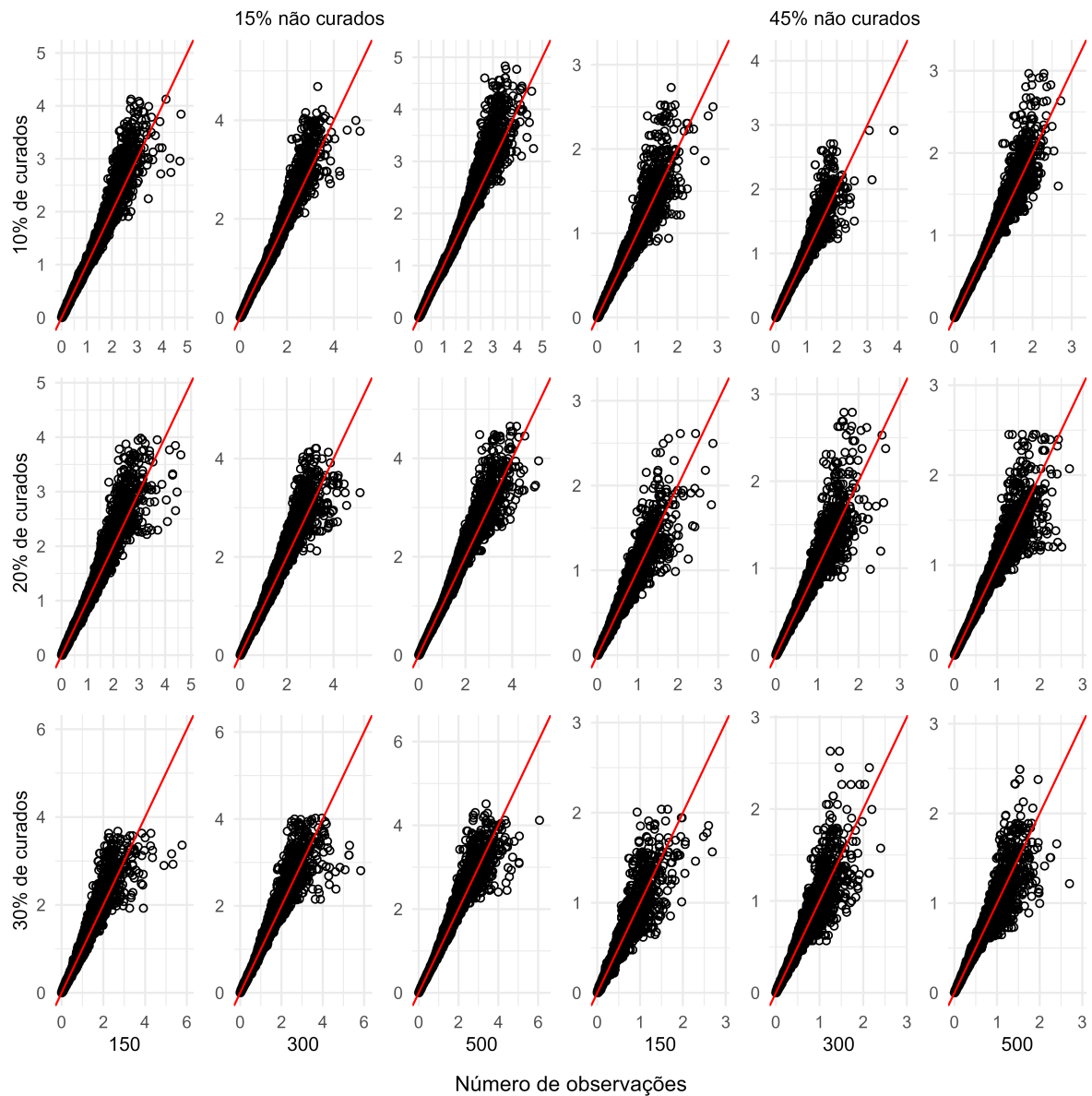
aqui, o mesmo acontece quando se estima o modelo com esquema de primeira ativação para dados gerados com um esquema de ativação geométrica com $\xi = 0, 1$.

Figura 25 – Resíduo observado *versus* resíduo esperado para a especificação errada no esquema de ativação aleatória com distribuição log-normal.



Fonte: o autor

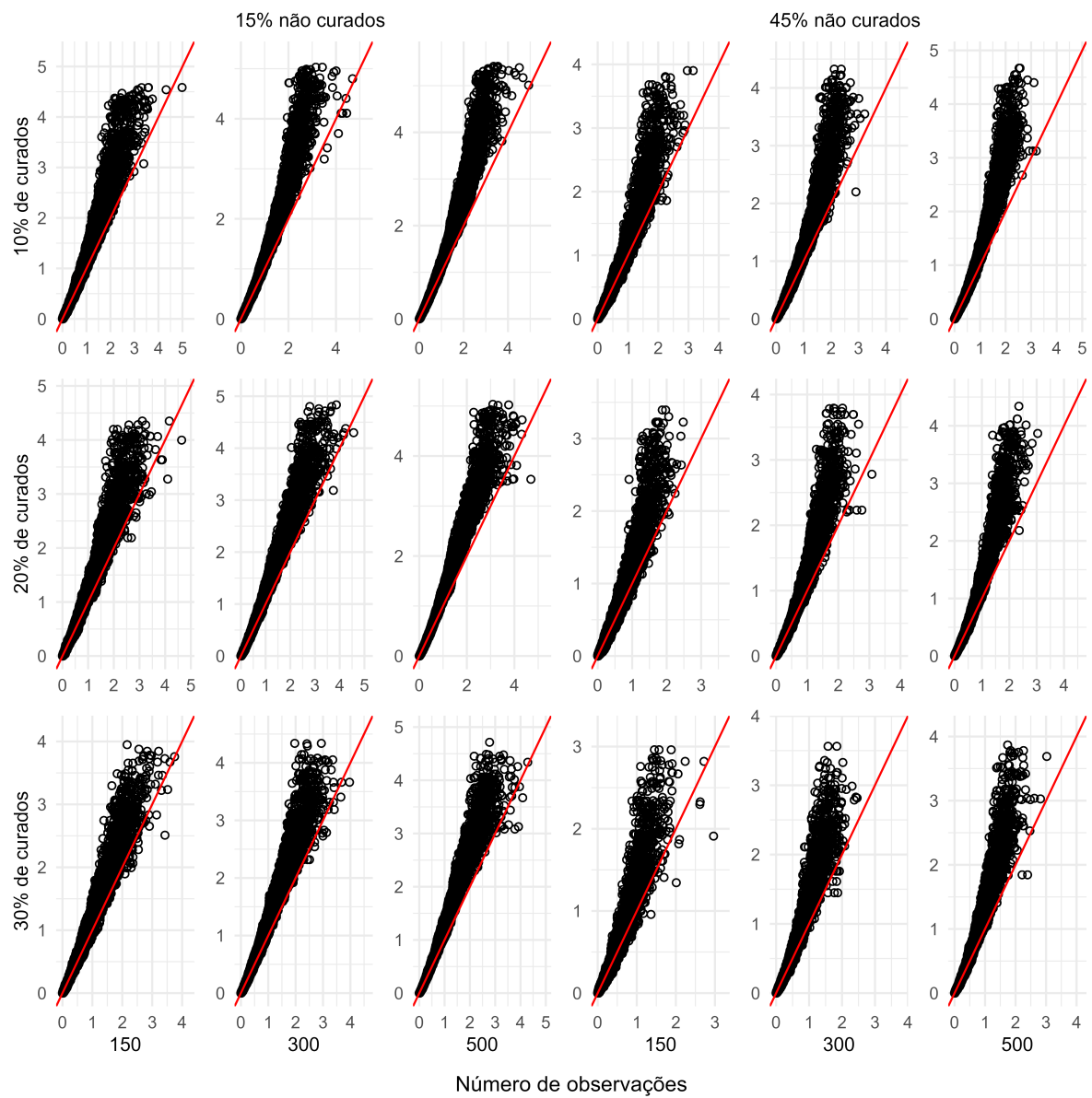
Figura 26 – Resíduo observado *versus* resíduo esperado para a especificação errada no esquema de ativação aleatória com distribuição log-t-Student com três graus de liberdade.



Fonte: o autor

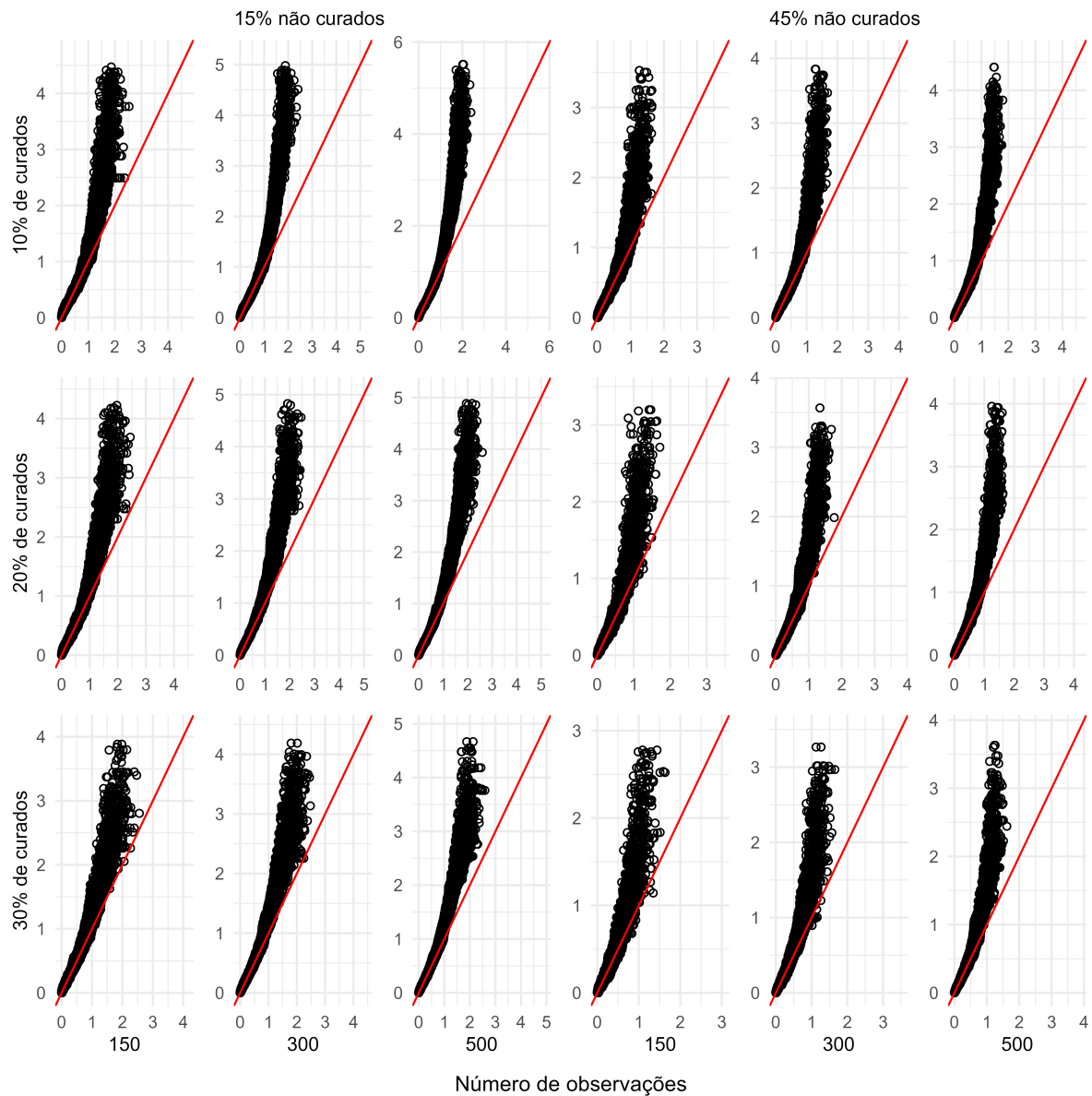
As Figuras 25 e 26 ilustram os gráficos gerados a partir de dados simulados sob o esquema de ativação aleatória, mas estimados considerando o esquema de primeira ativação. Observa-se uma inclinação distinta de 1, perceptível mesmo em resíduos intermediários, enquanto que na especificação correta esse comportamento se manifestava apenas nos resíduos mais altos, sendo que essa inclinação se acentua à medida que os resíduos aumentam. Além disso, verifica-se uma maior dispersão dos pontos com o aumento da censura, um padrão que se torna ainda mais evidente em amostras de maior tamanho amostral.

Figura 27 – Resíduo observado *versus* resíduo esperado para a especificação errada no esquema de ativação geométrica com distribuição log-normal.



Fonte: o autor

Figura 28 – Resíduo observado *versus* resíduo esperado para a especificação errada no esquema de ativação geométrica com distribuição log-t-Student com três graus de liberdade.



Fonte: o autor

Por fim, as Figuras 27 e 28 exibem os gráficos com dados simulados sob o esquema de ativação geométrica e modelo ajustado com o esquema de ativação aleatória. Nota-se, de forma muito similar ao caso da especificação incorreta da incidência, uma inclinação diferente de 1, que se acentua à medida que os resíduos aumentam, sendo mais pronunciada na distribuição log-t-Student com três graus de liberdade. Esse padrão torna-se mais evidente com o crescimento da amostra e com o aumento da $pcNC$, embora seu efeito tende a diminuir um pouco à medida que a fc se eleva.

4.2.6 Conclusão

Ao analisar as diversas simulações sob diferentes configurações de fc , $pcNC$, tamanhos amostrais e Cenários de especificação incorreta, constata-se que o resíduo de Cox–Snell apresenta desempenho variável: em alguns casos, o método é eficaz para detectar especificações incorretas, como na especificação incorreta da incidência do modelo ou do esquema no esquema de ativação aleatória estimada pela primeira ativação, ou de ativação geométrica estimada pela ativação aleatória; em outros, é possível identificar a especificação errada, mas com um desvio sutil em relação ao esperado, como ocorre na especificação incorreta da latência. No entanto, em certas situações, mostra-se pouco sensível, como nos casos de especificação incorreta do esquema de primeira ativação estimado pela ativação geométrica. Na prática, mesmo quando o resíduo indica a presença de especificação incorreta, ele não permite identificar com precisão a origem da especificação errada, já que desvios semelhantes podem ocorrer em diferentes tipos de especificação errada, como no caso do esquema de ativação geométrica com especificação errada de incidência e de esquema.

Os esquemas de primeira ativação e ativação geométrica apresentaram resultados semelhantes, com a distribuição log-t-Student com três graus de liberdade mostrando melhor desempenho, assim como observado na seção anterior. No entanto, essa semelhança também contribui para a limitação dos resíduos de Cox–Snell em distinguir entre os esquemas quando há especificação incorreta no esquema. Quando há especificação incorreta na distribuição de incidência, ambos os esquemas mostram desvios em relação ao comportamento esperado, os quais se intensificam com o aumento do tamanho amostral e da $pcNC$, embora esse efeito seja amenizado com o aumento da censura. De forma análoga, os resíduos também detectam a especificação incorreta da latência, mas essa detecção se torna mais difícil à medida que cresce a censura.

Por fim, o esquema de ativação aleatória já apresenta certo desvio de inclinação mesmo sob especificação correta e esse desvio se acentua na presença de erro na distribuição de latência. Nos casos de especificação incorreta da incidência, observa-se um número maior de pontos com comportamento semelhante ao Cenário corretamente especificado, ou seja, com um certo desvio na inclinação para os resíduos maiores, sendo esse efeito intensificado com o aumento da fc . Quando a especificação incorreta está no esquema, os resíduos tornam-se visivelmente mais dispersos do que nos casos de especificação correta, inclusive para valores intermediários, esse padrão se agrava à medida que aumentam a censura e o tamanho amostral.

Na prática, os resíduos de Cox–Snell funcionam como uma ferramenta de diagnóstico capaz de indicar a presença de má especificação, mas sem identificar sua origem exata. Eles se mostraram eficazes na detecção de erros na incidência ou no esquema, embora não sejam recomendados para distinguir entre primeira ativação e ativação geométrica. No caso de especificação incorreta da latência, conseguem apontar desvios, porém de forma sutil, não sendo indicados em cenários com altas taxas de censura. O esquema de ativação aleatória, por sua vez, já apresenta desvios mesmo sob especificação correta, que se acentuam quando há especificação incorreta, o que dificulta avaliar se o desvio observado é suficientemente grande para caracterizar um erro de modelagem ou apenas o desvio esperado do próprio esquema. Dessa forma, o uso dos resíduos de Cox–Snell não é recomendado para o esquema de ativação aleatória.

5 APLICAÇÕES

5.1 APLICAÇÃO 1- ESTUDO CLÍNICO E1684

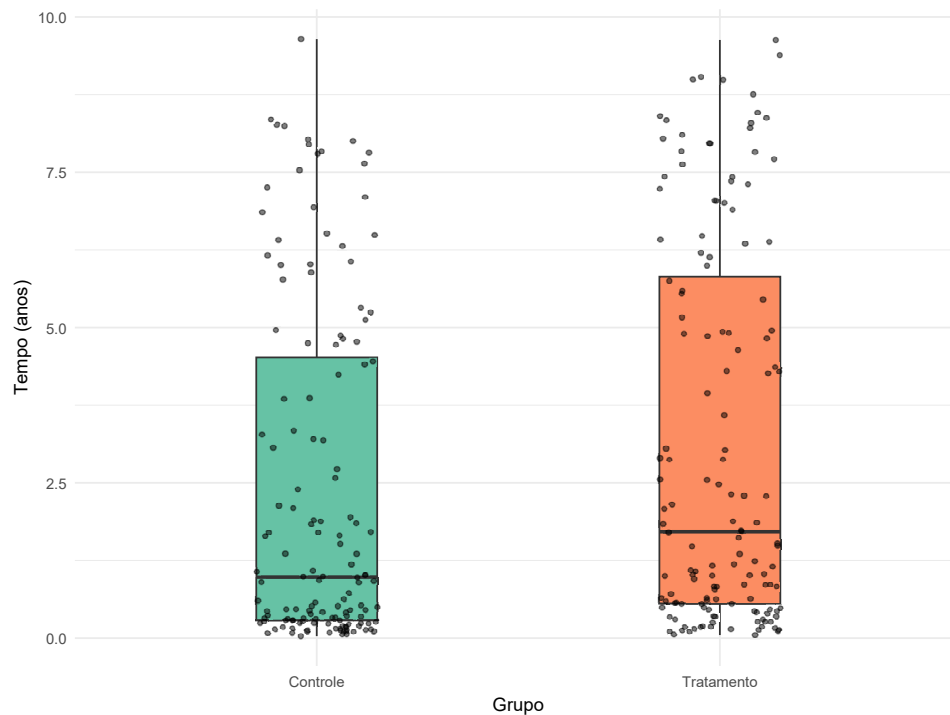
Consideramos os dados de uma amostra com 284 observações, correspondentes a pacientes acompanhados pelo Eastern Cooperative Oncology Group (ECOG) no ensaio clínico de fase III denominado E1684, realizado entre 1984 e 1993, com pacientes entrando no estudo até 1990. O estudo foi desenhado para avaliar o efeito da administração de interferon alfa-2b em alta dose como terapia adjuvante para pacientes com melanoma de alto risco. Os resultados do estudo foram publicados por Kirkwood et al. (1996), em que, por meio do teste log-rank foi detectada diferença significativa entre as curvas de sobrevivência dos grupos de tratamento. Além disso, utilizando o modelo de riscos proporcionais de Cox foi identificado um efeito significativo positivo do tratamento com interferon na sobrevida livre de recidiva, ou seja, no tempo até o paciente apresentar recidiva ou óbito. No entanto, não foi proposto um modelo que explicasse a probabilidade de cura dos pacientes e sua relação com o tratamento, o que pode ser feito por meio dos modelos de fração de cura. Em seu trabalho, Corbiere e Joly (2007) analisaram esses dados utilizando dois modelos de fração de cura: um em que a latência segue a distribuição Weibull e outro com estrutura de riscos proporcionais de Cox para a latência. Contudo, ambos os modelos consideraram apenas a estrutura do modelo de mistura padrão. Nesse sentido, o modelo com estrutura de esquemas de ativação, que inclui o modelo de mistura como caso particular, pode ser mais adequado. As variáveis observadas para cada paciente foram as seguintes: Tratamento (grupo de tratamento, sendo 0 para grupo controle e 1 para grupo tratado com interferon), Tempo (tempo de sobrevivência livre de recidiva observado, em anos), Status (indicador de censura, sendo 0 para censurado e 1 para ocorrência de morte ou recidiva), Idade (idade do paciente no momento da divisão dos grupos tratamento e controle, em anos), Sexo (sexo do paciente, sendo 0 para masculino e 1 para feminino).

Tabela 17 – Proporção de censura por grupo de tratamento

Tratamento	Status = 0	Status = 1
Controle	0,2500	0,7500
Tratado	0,3681	0,6319

Fonte: o autor

Figura 29 – Gráficos box-plot do tempo por grupo de tratamento.



Fonte: o autor

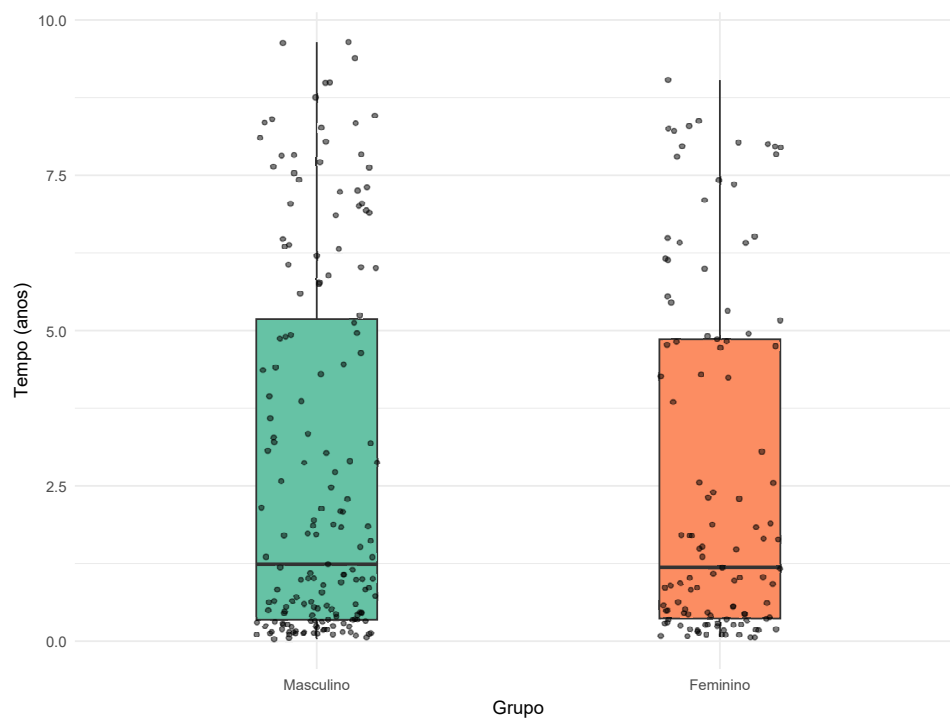
A Tabela 17 mostra que a proporção de indivíduos que morreram ou apresentaram recidiva é aproximadamente 12% menor no grupo tratado com interferon, o que pode indicar uma eficácia do tratamento na redução desses desfechos. Na Figura 29, observa-se uma assimetria à direita moderada, sendo aproximadamente 1 para o grupo controle e menor para o grupo tratamento com aproximadamente 0,69. Em relação à mediana, temos que 50% dos eventos (morte ou recidiva) ocorrem em menos de 1 ano no grupo controle, enquanto no grupo tratado ocorrem em menos de 1 ano e 9 meses. Além disso, há pacientes com longos tempos de sobrevivência em ambos os grupos, o que pode indicar a presença de indivíduos potencialmente curados. No entanto, o grupo tratado apresenta maior concentração de tempos elevados de sobrevivência, com os quartis de 25%, 50% e 75% mais altos em comparação ao grupo controle.

Tabela 18 – Proporção de censura por sexo

Sexo	Status = 0	Status = 1
Masculino	0,3041	0,6959
Feminino	0,3186	0,6814

Fonte: o autor

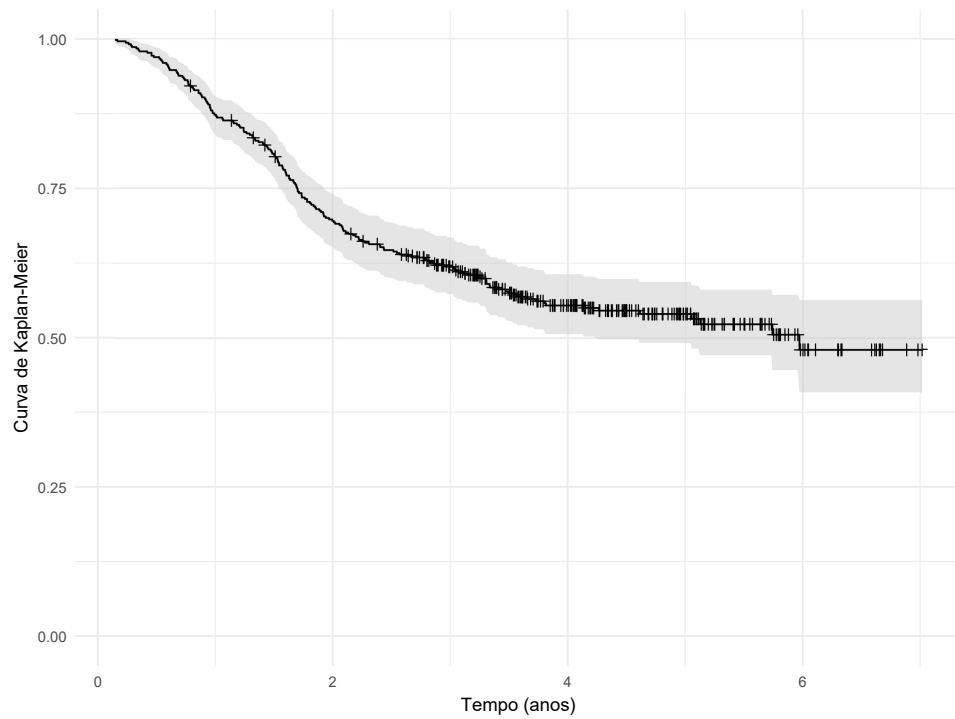
Figura 30 – Gráficos box-plot do tempo por sexo.



Fonte: o autor

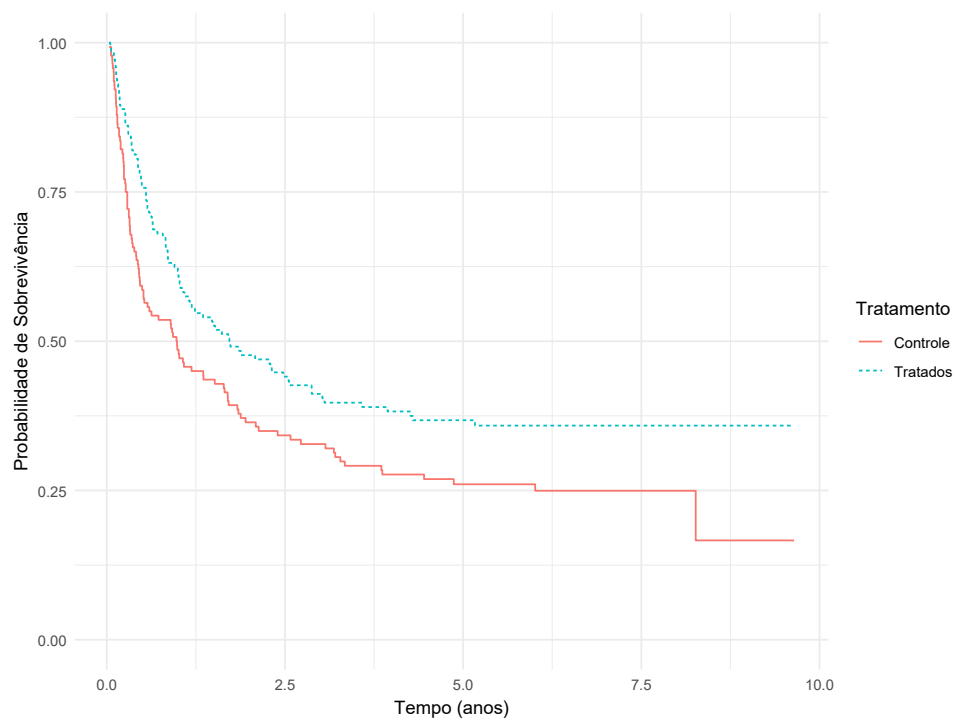
A Tabela 18 mostra que não há uma diferença expressiva na proporção de indivíduos que morreram ou apresentaram recidiva em relação ao sexo. Observa-se uma leve predominância de eventos no grupo masculino. A Figura 30 também evidencia pouca variação nos tempos de sobrevivência entre os grupos, com quartis bastante próximos e distribuição assimétrica à direita moderada em ambos os casos, com aproximadamente 0,83 para o grupo masculino e 0,84 para o feminino.

Figura 31 – Gráfico Kaplan-Meier.



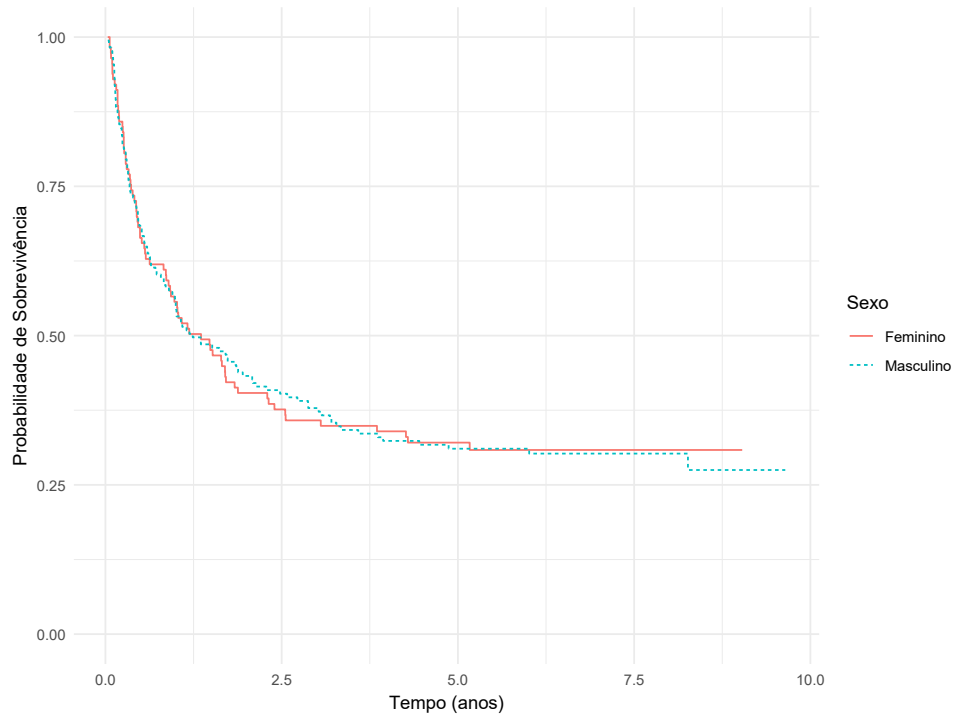
Fonte: o autor

Figura 32 – Gráfico Kaplan-Meier por tratamento.



Fonte: o autor

Figura 33 – Gráfico Kaplan-Meier por sexo.



Fonte: o autor

Da Figura 31, temos a curva de Kaplan-Meier com seu intervalo de confiança de 95% e seus indicadores de censura, observa-se que a curva apresenta um platô por mais de cinco anos. Considerando que o tempo de acompanhamento do estudo é de quase dez anos, a presença de indivíduos curados pode ser responsável por esse platô durar tanto. A Figura 32 do gráfico do EKM por tratamento apresentam o platô por quase 5 anos apenas no grupo de tratamento, enquanto o grupo controle apresenta platô por pouco mais de 1 ano. Já o gráfico do EKM por sexo, apresentado na Figura 33 mostra curvas semelhantes entre os estratos, com o grupo feminino apresentando um platô durante mais tempo que o grupo masculino.

Seja T_{ijk} o tempo em anos de sobrevivência sem recidiva para o i -ésimo indivíduo com j -ésimo tratamento e k -ésimo sexo e supondo um modelo com esquema de ativação R_{ijk} , incidência M_{ijk} , latência $\tilde{Y}_{ijk}$ e com parte sistemática dada em θ_{ijk} , temos que:

$$S_{pop}(t_{ijk}) = P_{\theta_{ijk}}(R_{ijk} > M_{ijk}) \\ + E_{R_{ijk}, M_{ijk}} \left[IB(S_{\tilde{Y}_{ijk}}; M_{ijk} - R_{ijk} + 1, R_{ijk}) I(M_{ijk} \geq R_{ijk} \geq 1) \right]$$

e

$$b^{-1}(\theta_{ijk}) = \beta_0 + \beta_1 \text{idade}_{ijk} + \gamma_j + \alpha_k, \quad j, k = 1, 2, i = 1 \dots 284,$$

em que β_0 é o intercepto, β_1 o efeito da $idade_{ijk}$, com $idade_{ijk}$ denotando a idade do i -ésimo indivíduo com j -ésimo tratamento e k -ésimo sexo, γ_j é o efeito do j -ésimo tratamento, α_k é o efeito do k -ésimo sexo e $b(\cdot)$ é a função de ligação. Assume-se também que $\gamma_1 = 0$ e $\alpha_1 = 0$ como casela de referência.

Para modelagem dos dados seguindo o modelo acima foi considerado as distribuições de incidência Bernoulli($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_{ijk})$), Poisson($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_{ijk})$) e geométrica($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_{ijk})$), as distribuições da latência foram as distribuições log-normal($\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim} LN(\eta, \phi)$) e log-t-Student($\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim} LT(\eta, \phi)$) com graus de liberdade 3, 4, 5 e 6 e os esquemas de primeira ativação($R_{ijk} = 1$), de ativação aleatória($R_{ijk} \stackrel{\text{ind}}{\sim} U\{1 \dots M_{ijk}\}$) e ativação geométrica($R_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\xi)$). A estimação dos parâmetros foi feita pelo método de estimação por máxima verossimilhança utilizando a verossimilhança marginal e a escolha do modelo foi feita pelos critérios de menor **AIC** e **BIC**.

Ressalta-se que os modelos com esquema de primeira ativação são casos particulares (ou encaixados) dos modelos com ativação geométrica, quando $\xi = 1$. Nesse contexto, é possível realizar um **TRV** para avaliar a necessidade de utilizar o modelo mais complexo (com ativação geométrica) ou o modelo mais simples (com primeira ativação). Os resultados do **TRV** para os diferentes modelos considerados com diferentes incidências e latências mostram que não há evidências suficientes para rejeitar a hipótese nula $H_0: \xi = 1$ com todos os testes apresentando valor-p superior a 0,96. Portanto, apenas os modelos com primeira ativação e ativação aleatória foram considerados.

Tabela 19 – Modelos ajustados com diferentes esquemas de ativação

Esquema (R)	Incidência (M)	Latência (Y)	AIC	BIC
Aleatória	Geométrica	log-t-Student ($\nu = 3$)	370,3987	392,2926
Primeira	Bernoulli	log-t-Student ($\nu = 3$)	370,3988	392,2926
Aleatória	Bernoulli	log-t-Student ($\nu = 3$)	370,3988	392,2926
Aleatória	Poisson	log-t-Student ($\nu = 3$)	370,4277	392,3216
Aleatória	Geométrica	log-t-Student ($\nu = 4$)	373,1760	395,0698
Primeira	Bernoulli	log-t-Student ($\nu = 4$)	373,1760	395,0698
Aleatória	Bernoulli	log-t-Student ($\nu = 4$)	373,1760	395,0698
Aleatória	Poisson	log-t-Student ($\nu = 4$)	373,2037	395,0976
Aleatória	Geométrica	log-t-Student ($\nu = 5$)	375,0006	396,8944
Primeira	Bernoulli	log-t-Student ($\nu = 5$)	375,0006	396,8944
Aleatória	Bernoulli	log-t-Student ($\nu = 5$)	375,0006	396,8944
Aleatória	Poisson	log-t-Student ($\nu = 5$)	375,0277	396,9215
Primeira	Poisson	log-t-Student ($\nu = 3$)	375,0805	396,9744
Primeira	Poisson	log-t-Student ($\nu = 4$)	375,8685	397,7624
Primeira	Bernoulli	log-t-Student ($\nu = 6$)	376,2918	398,1856
Aleatória	Bernoulli	log-t-Student ($\nu = 6$)	376,2918	398,1856
Aleatória	Geométrica	log-t-Student ($\nu = 6$)	376,2920	398,1859
Aleatória	Poisson	log-t-Student ($\nu = 6$)	376,3184	398,2123
Primeira	Geométrica	log-t-Student ($\nu = 5$)	376,5965	398,4903
Primeira	Poisson	log-t-Student ($\nu = 5$)	376,6095	398,5033
Primeira	Geométrica	log-t-Student ($\nu = 6$)	376,6560	398,5499
Primeira	Geométrica	log-t-Student ($\nu = 4$)	376,8780	398,7719
Primeira	Poisson	log-t-Student ($\nu = 6$)	377,2368	399,1307
Primeira	Geométrica	log-t-Student ($\nu = 3$)	378,1575	400,0514
Primeira	Geométrica	log-normal	380,9687	402,8625
Primeira	Poisson	log-normal	382,5790	404,4728
Aleatória	Geométrica	log-normal	383,9203	405,8141
Primeira	Bernoulli	log-normal	383,9205	405,8143
Aleatória	Bernoulli	log-normal	383,9205	405,8143
Aleatória	Poisson	log-normal	383,9448	405,8387

Fonte: o autor

A Tabela 19 apresenta os valores de AIC e BIC para todos os modelos dos esquemas de primeira ativação e ativação aleatória, ordenados pelo menor AIC. Por apresentarem os menores valores em ambas as métricas, os modelos de primeira ativação com incidência Bernoulli, ativação aleatória com incidência Geométrica e Poisson, todos com latência log-t-Student com 3 graus de liberdade, podem ser escolhidos, sendo a escolha entre eles fundamentada na

interpretação técnica do mecanismo biológico que leva à recidiva ou morte por câncer. Como os três modelos apresentam resultados equivalentes e devido à ausência de interpretação técnica específica no presente estudo, foi considerado o modelo com ativação aleatória e incidência geométrica.

Define-se então, um modelo com esquema de ativação dado por $R_{ijk} \stackrel{\text{ind}}{\sim} U\{1 \dots M_{ijk}\}$, no qual a ocorrência de morte ou recidiva tem igual probabilidade de ocorrer por qualquer uma das causas da incidência M_{ijk} . A variável de incidência segue uma distribuição geométrica, $M_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_{ijk})$, enquanto a latência é modelada por $\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim} LT(3)$. A parte sistemática está associada ao parâmetro θ_{ijk} , e a **FSP** é dada por

$$S_{pop}(t_{ijk}) = \theta_{ijk} + (1 - \theta_{ijk})S_{\tilde{Y}_i}(t_{ijk}),$$

em que

$$\log\left(\frac{\theta_{ijk}}{1 - \theta_{ijk}}\right) = \beta_0 + \beta_1 \text{idade}_{ijk} + \gamma_j + \alpha_k, \quad j, k = 1, 2, i = 1 \dots 284.$$

Tabela 20 – Estimativas dos coeficientes do modelo em investigação

Coeficiente	Estimativa	EP	Z	Valor-p	IC 2,5%	IC 97,5%
η	0,6593	0,0699	9,4279	< 0,001	0,5222	0,7963
ϕ	1,2344	0,1923	6,4182	< 0,001	0,8574	1,6114
β_0 (Intercepto)	-1,4723	0,2966	-4,9635	< 0,001	-2,0536	-0,8909
γ_1 (tratamento)	0,6930	0,3262	2,1244	0,034	0,0536	1,3323
β_1 (idade)	-0,0167	0,0125	-1,3406	0,180	-0,0411	0,0077
α_1 (Feminino)	0,0323	0,3224	0,1002	0,920	-0,5996	0,6643

Fonte: o autor

Tabela 21 – Estimativas dos coeficientes do modelo final

Coeficiente	Estimativa	EP	Z	Valor-p	IC 2,5%	IC 97,5%
η	0,6581	0,0696	9,4546	< 0,001	0,5216	0,7945
ϕ	1,2308	0,1914	6,4295	< 0,001	0,8556	1,6060
β_0 (Intercepto)	-1,4290	0,2629	-5,4349	< 0,001	-1,9444	-0,9137
γ_1 (tratamento)	0,6638	0,3232	2,0538	0,040	0,0303	1,2973

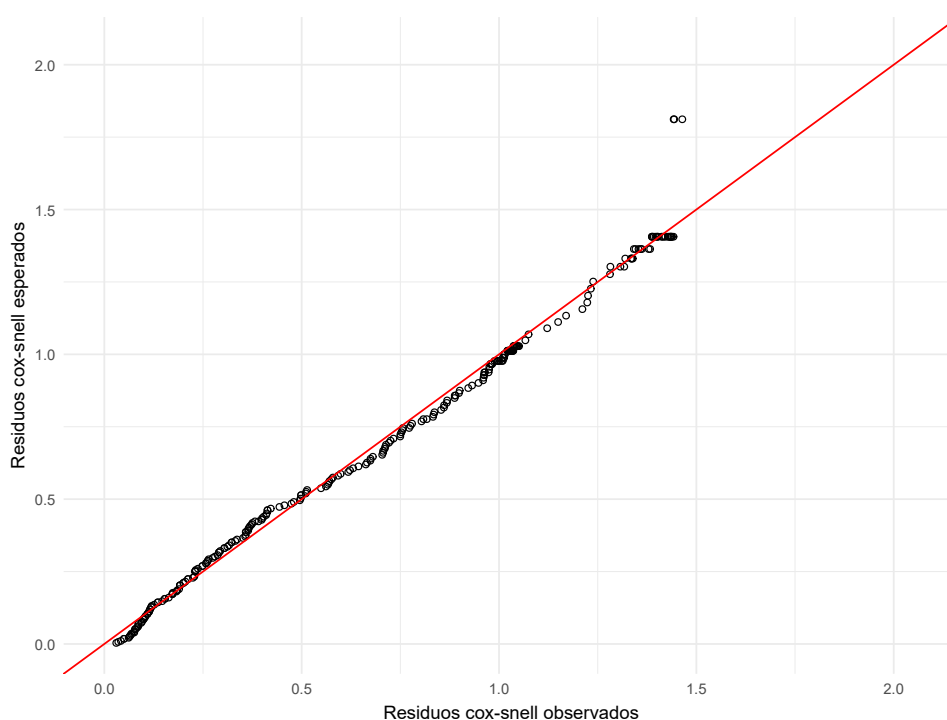
Fonte: o autor

A Tabela **20** apresenta os resultados do ajuste do modelo com todas as covariáveis. Verifica-se que, ao nível de 5% de significância, apenas a variável tratamento é estatisticamente significativa. Utilizando o método *stepwise* para seleção de modelos com base nos critérios **AIC**

e **BIC**, observa-se que o modelo contendo apenas a covariável tratamento apresenta o menor **AIC**, enquanto o modelo com apenas o intercepto é aquele que possui o menor **BIC**. Aplicando o **TRV** para comparar o modelo com apenas o intercepto com aquele que inclui a covariável tratamento, obtém-se um valor- p de 0,0344, o que permite rejeitar o modelo mais simples.

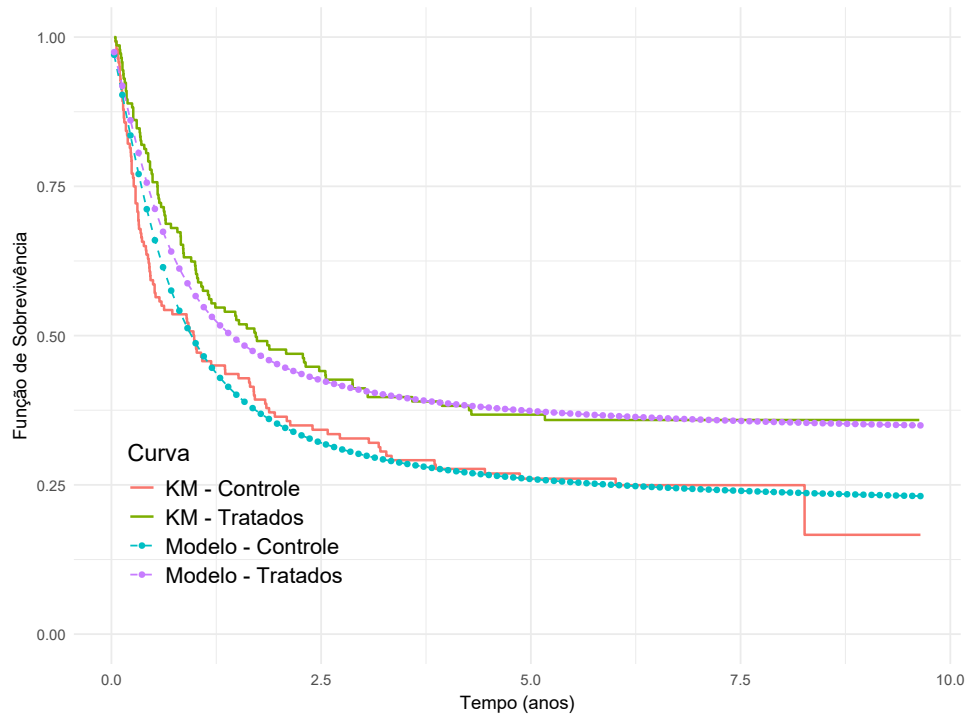
Dessa forma, o modelo final escolhido foi aquele que inclui a covariável tratamento e o intercepto. Os resultados desse ajuste estão apresentados na Tabela **21** em que todos os parâmetros se mostram significativos ao nível de 5%.

Figura 34 – Gráfico dos resíduos de Cox-Snell.



Fonte: o autor

Figura 35 – Gráfico Kaplan-Meier por categoria de nódulos.



Fonte: o autor

A Figura 34 apresenta os resíduos de Cox–Snell observados em comparação com os valores esperados. Nota-se que o modelo está bem especificado, pois a maioria dos pontos está sobre (ou próximo) à linha de referência, com apenas alguns casos mais afastados, os quais podem indicar possíveis observações discrepantes.

A Figura 35 exibe as curvas do EKM para cada categoria de tratamento, comparadas com as curvas estimadas pelo modelo. Observa-se que o ajuste é satisfatório, ao apresentar curvas estimadas próximas das curvas do EKM. Especialmente nos anos mais recentes as curvas ficam muito próximas, o que indica uma boa modelagem da fração de cura, sobretudo na categoria que recebeu o tratamento.

Assim, o modelo final ajustado é dado por

$$\hat{S}_{pop}(t_{ij}) = \hat{\theta}_j + (1 - \hat{\theta}_j)S_{\tilde{Y}_j}(t_{ij})$$

e

$$\log \left(\frac{\hat{\theta}_j}{1 - \hat{\theta}_j} \right) = \beta_0 + \gamma_j, \quad j = 1, 2, i = 1 \dots 284.$$

Esse modelo apresenta as seguintes probabilidades de cura, a depender do tratamento:

$$\hat{P}(\text{"Curado"} \mid \text{tratamento} = \text{controle}) = \hat{\theta}_1 = 0,1932,$$

$$\hat{P}(\text{"Curado"} \mid \text{tratamento} = \text{tratado}) = \hat{\theta}_2 = 0,3175.$$

Essas probabilidades indicam que o tratamento é eficaz no sentido de aumentar as chances de cura dos pacientes, resultado ainda mais evidente pela razão de chances $\hat{\psi}_\gamma = 1,9421$. Mostrando que um indivíduo submetido ao tratamento apresenta aproximadamente 94,2% mais chances de estar curado em comparação com aquele que não recebeu o tratamento. Além disso, o intervalo de confiança de 95% para a razão de chances é de 1,0322 a 3,6542 e como o valor 1 não está contido nesse intervalo, conclui-se que o tratamento é significativo para a sobrevivência livre de recidiva.

5.2 APLICAÇÃO 2- ESTUDO CLÍNICO E1690

Utilizamos os dados de uma amostra com 426 observações, provenientes do ensaio clínico de fase III também da ECOG, denominado de E1690 e publicados por [Kirkwood et al. \(2000\)](#). O estudo foi desenvolvido entre 1991 e 1998, com pacientes entrando no estudo até 1995 e teve como objetivo confirmar os achados do ensaio E1684, além de investigar o efeito de doses mais baixas de interferon alfa-2b na sobrevida de pacientes com melanoma de alto risco, devido aos efeitos colaterais consideráveis das altas doses. Os pacientes incluídos apresentavam as mesmas características clínicas do estudo anterior e a análise estatística foi realizada utilizando o teste log-rank ponderado, sob a suposição de taxas de cura diferentes entre as curvas. Não foi detectada diferença significativa entre os grupos de tratamento e controle. Além disso, utilizando o modelo de riscos proporcionais de Cox, foi identificado que o tratamento com dosagem baixa não foi significativo na sobrevida dos indivíduos. Embora no estudo tenha sido utilizado um teste de hipótese que leva em consideração a fração de curados, a probabilidade de cura ainda não foi analisada. [Cooner et al. \(2007\)](#) realizaram a análise modelando a fração de cura por meio do modelo com esquemas de ativação desenvolvido no próprio artigo, sob a abordagem bayesiana. Neste trabalho abordaremos o problema utilizando o modelo de fração de cura log-simétrico no enfoque clássico sob vários esquemas de ativação. As variáveis observadas para cada paciente são as seguintes: Tempo (tempo até a recidiva da doença, em anos), Status (indicador de censura, sendo 0 para censurado e 1 para ocorrência de morte ou recidiva), Tratamento (grupo de tratamento, sendo 0 para grupo controle e 1 para grupo tratado com interferon), Sexo (sexo do paciente, sendo 0 para masculino e 1 para feminino), Idade (idade do paciente no momento da divisão dos grupos tratamento e controle, em anos), Nódulo (Categoria do nódulo: 1 para nenhum nódulo, 2 para 1 nódulo, 3 para 2 a 3 nódulos e 4 para mais que 3 nódulos), P.S (status da performance no estudo, sendo 0 completamente ativo e

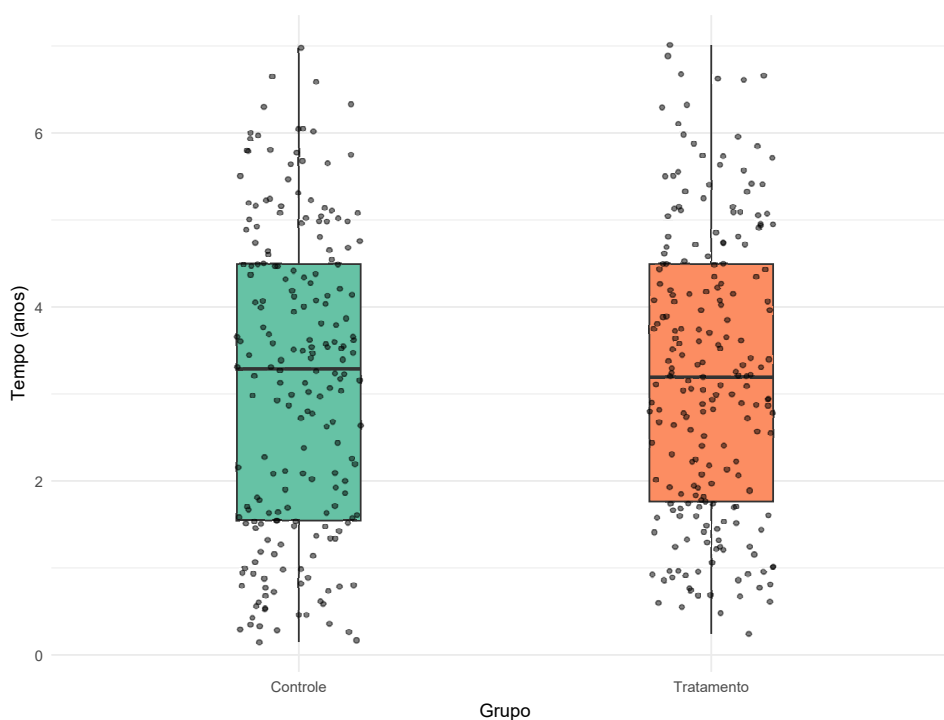
1 para outro), Breslow (Espessura do tumor).

Tabela 22 – Proporção de censura por grupo de tratamento

Tratamento	Status = 0	Status = 1
Controle	0,5686	0,4314
Tratado	0,5446	0,4554

Fonte: o autor

Figura 36 – Gráficos box-plot do tempo por grupo de tratamento.



Fonte: o autor

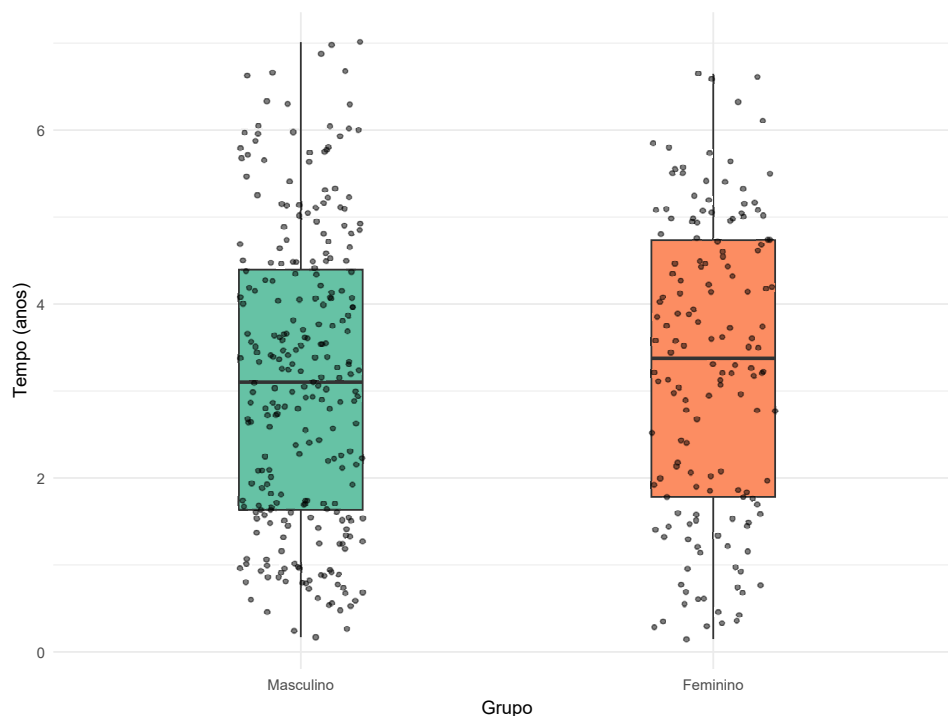
Vimos pela Tabela 22, em conjunto com a Figura 36, que uma redução na dose do tratamento pode ter comprometido sua eficácia, pois a proporção de mortes e os tempos até o óbito são bastante semelhantes entre os grupos. Observa-se também uma assimetria menos acentuada nesses dados, com uma assimetria quase inexistente de aproximadamente 0,03 no grupo controle e uma assimetria de aproximadamente 0,20 no grupo de tratados.

Tabela 23 – Proporção de censura por sexo

Sexo	Status = 0	Status = 1
Masculino	0,5437	0,4563
Feminino	0,5779	0,4221

Fonte: o autor

Figura 37 – Gráficos box-plot do tempo por sexo



Fonte: o autor

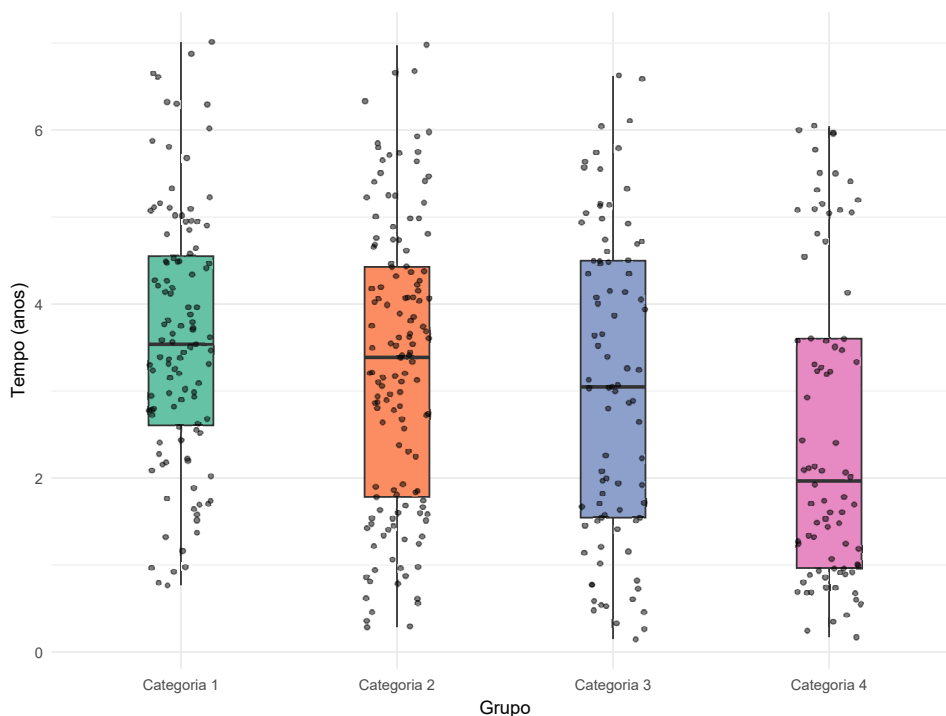
De forma semelhante ao estudo anterior, a Tabela 23 indica que a proporção de óbitos não difere substancialmente entre os sexos. Neste caso, também há uma ligeira predominância de eventos no grupo masculino. A Figura 30 apresenta tempos de sobrevivência similares entre os grupos, embora os quartis do grupo feminino sejam ligeiramente superiores. Além disso, nota-se baixa assimetria nas distribuições dos tempos para ambos os sexos.

Tabela 24 – Proporção de censura por categoria do nódulo

Nódulo	Status = 0	Status = 1
Categoria 1	0,6757	0,3243
Categoria 2	0,6131	0,3869
Categoria 3	0,5287	0,4713
Categoria 4	0,3293	0,6707

Fonte: o autor

Figura 38 – Gráficos box-plot do tempo por categoria do nódulo



Fonte: o autor

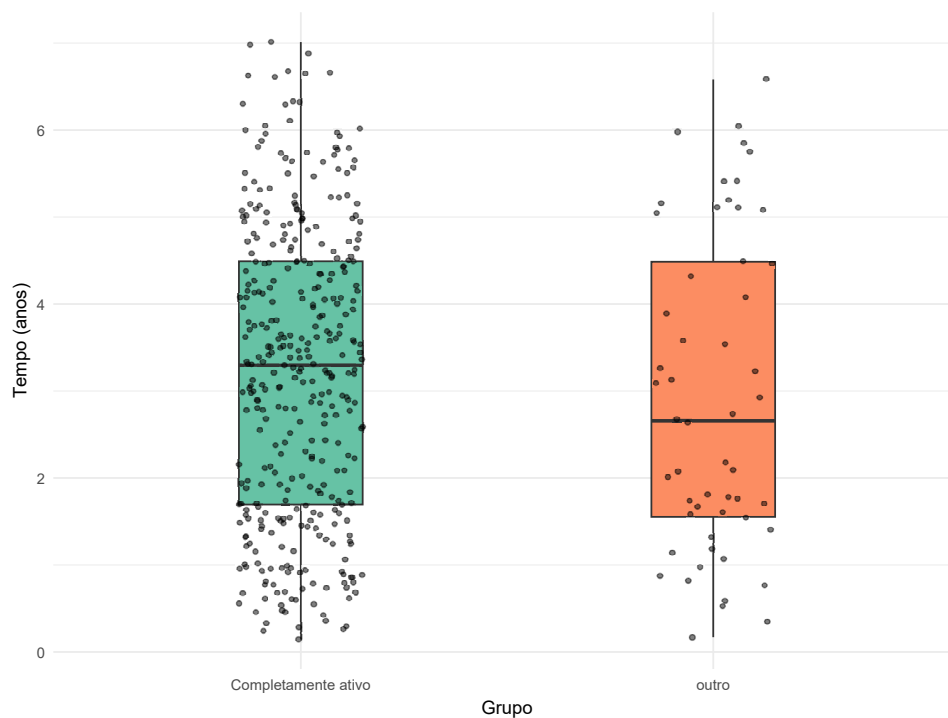
Na Tabela 24, observa-se indícios de que as categorias do nódulo influenciam a proporção de indivíduos que morreram: quanto maior a categoria do nódulo, maior a proporção de óbitos. Esse padrão também é evidenciado na Figura 38, onde, conforme a categoria do nódulo aumenta, os tempos até a morte diminuem, acompanhados da redução dos seus quartis. No caso da categoria 4, observa-se ainda uma assimetria à direita moderada de aproximadamente 0,56.

Tabela 25 – Proporção de censura por status de performance

P.S.	Status = 0	Status = 1
Completamente ativo	0,5647	0,4353
Outro	0,5000	0,5000

Fonte: o autor

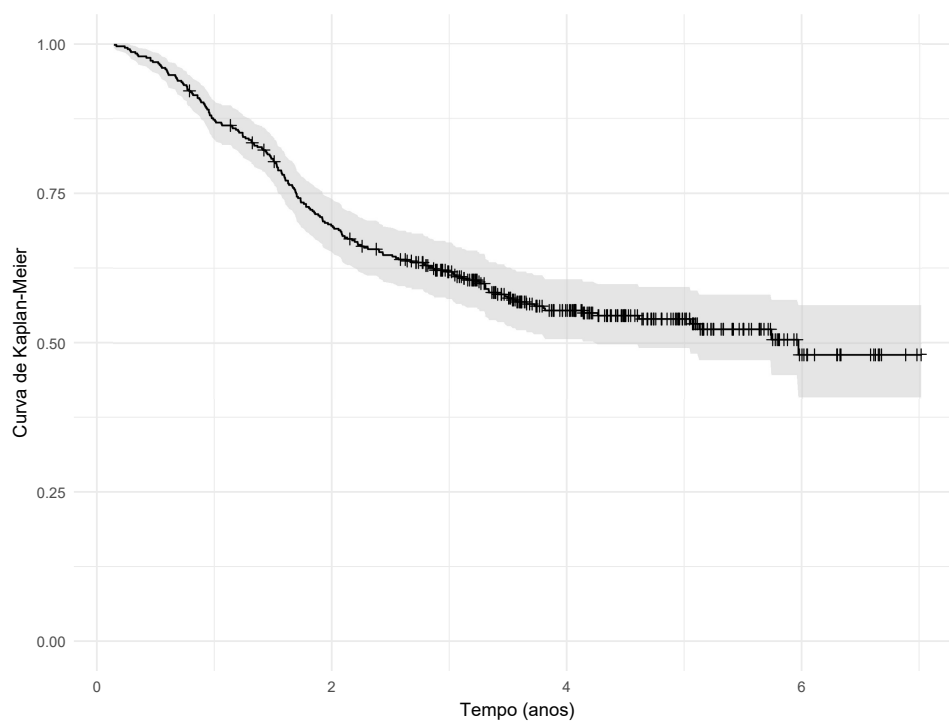
Figura 39 – Gráficos box-plot do tempo por status de performance



Fonte: o autor

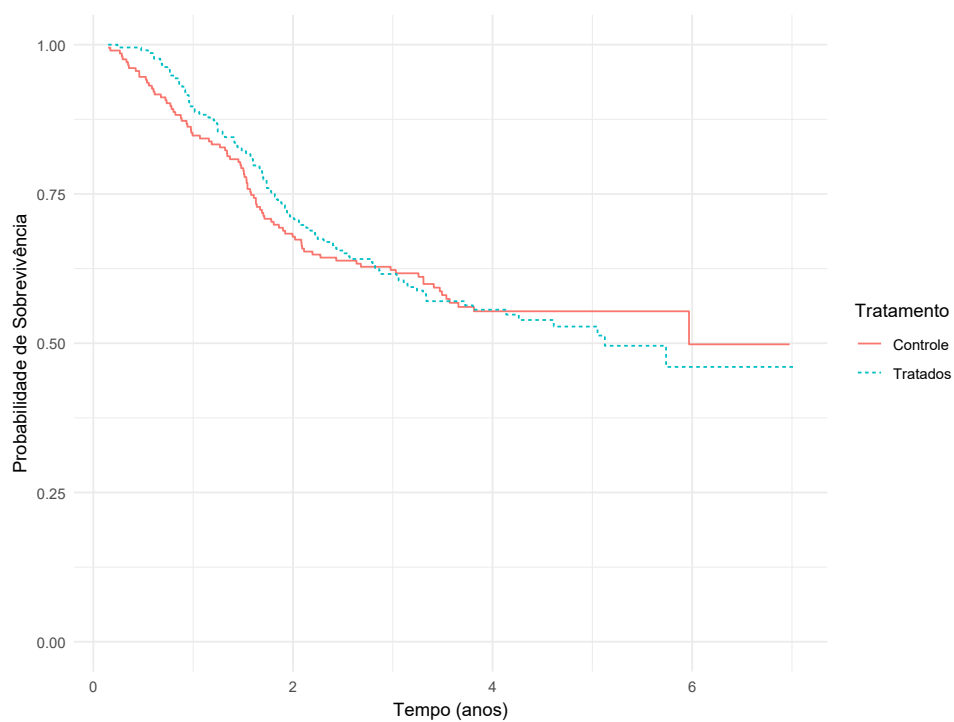
A Tabela [25](#) indica que o status de performance está associado a uma proporção ligeiramente maior de mortes entre os indivíduos que não estão completamente ativos. Já na Figura [39](#), os tempos dos indivíduos completamente ativos apresentam uma assimetria à esquerda, enquanto aqueles em outros status também apresentam assimetria, mas a direita.

Figura 40 – Gráfico Kaplan-Meier.



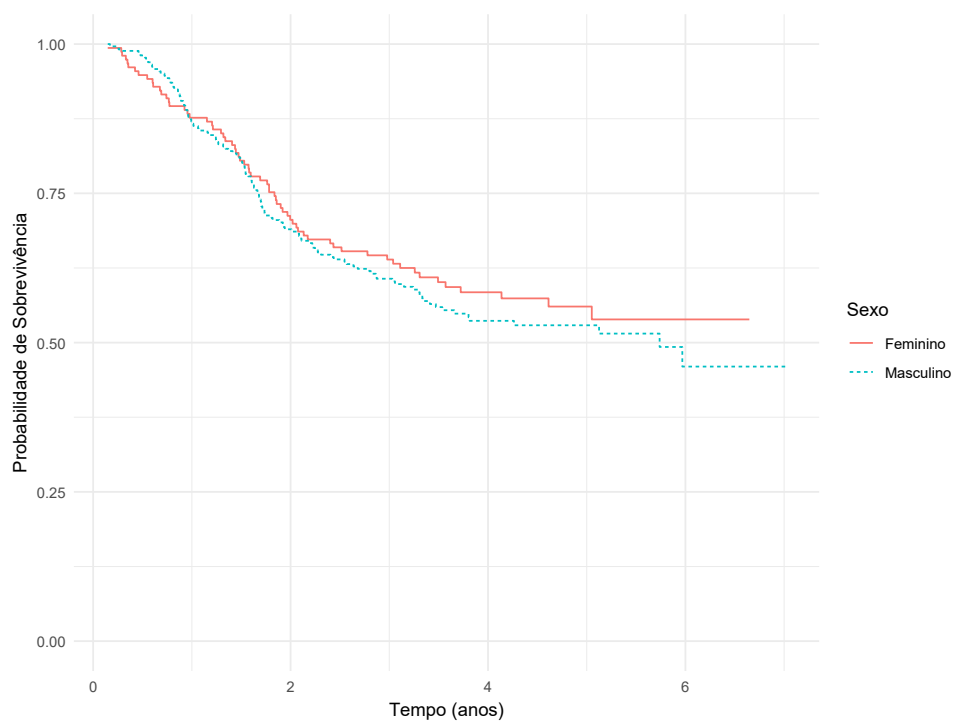
Fonte: o autor

Figura 41 – Gráfico Kaplan-Meier por tratamento.



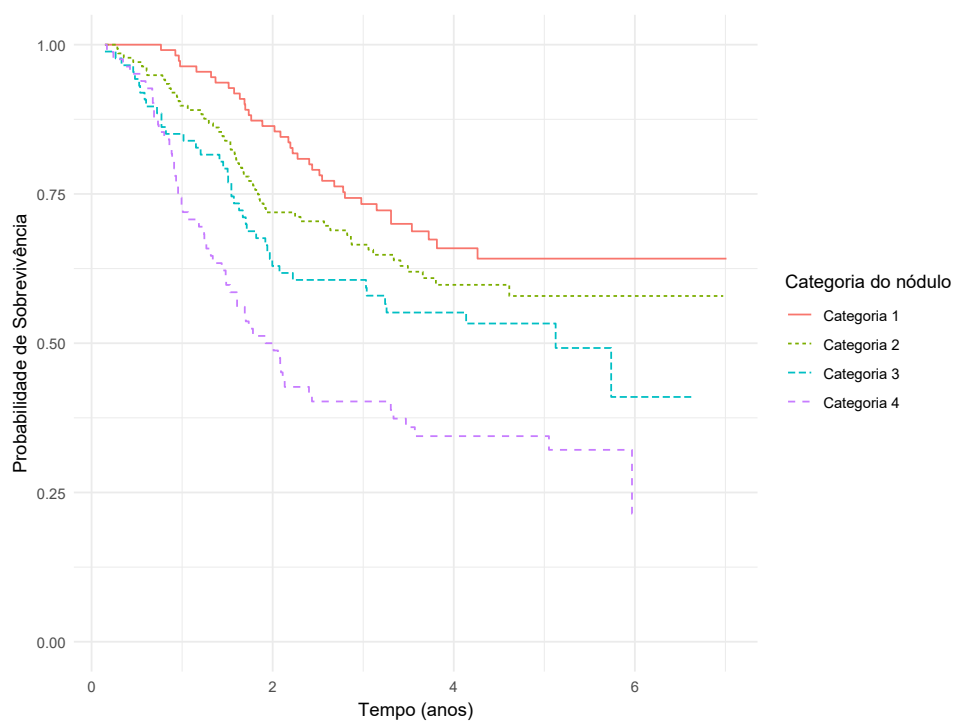
Fonte: o autor

Figura 42 – Gráfico Kaplan-Meier por sexo.



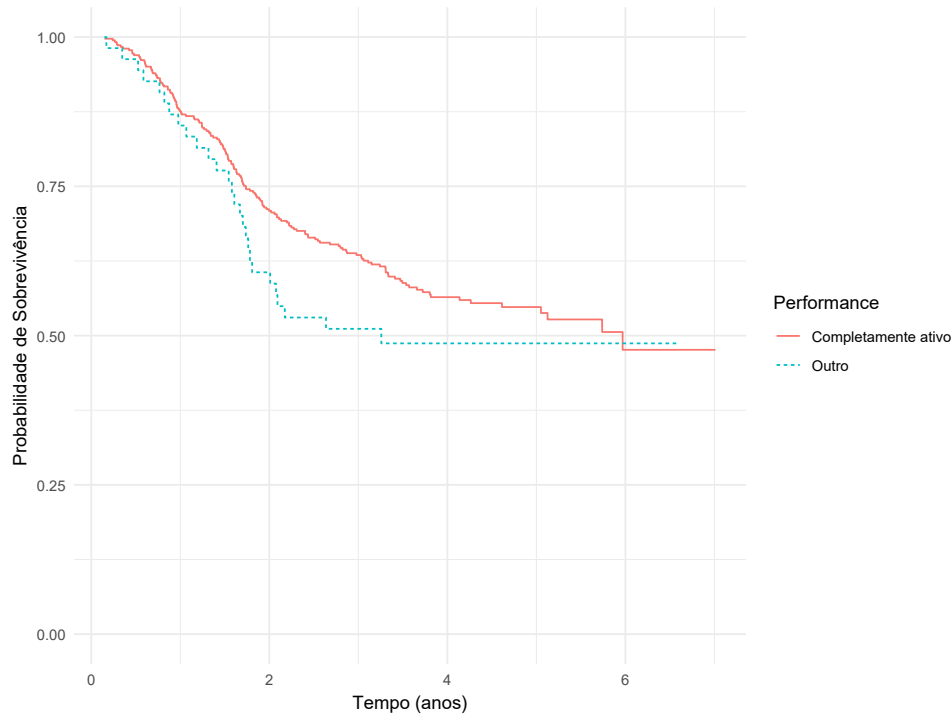
Fonte: o autor

Figura 43 – Gráfico Kaplan-Meier por categoria de nódulo.



Fonte: o autor

Figura 44 – Gráfico Kaplan-Meier por status de performance.



Fonte: o autor

Na Figura 40, a curva de Kaplan-Meier com seu intervalo de confiança e indicadores de censura, também exibe um platô, agora por mais de três anos. Com um tempo de acompanhamento de aproximadamente sete anos, a presença de indivíduos curados pode ser responsável por esse platô. A Figura 41 apresenta o EKM estratificado por tratamento, evidenciando curvas de sobrevivência semelhantes entre os grupos tratado e controle, sendo que o grupo controle começa o platô antes. Estratificando o EKM por sexo, a Figura 42 mostra que, assim como na aplicação anterior, os dois grupos apresentam curvas de sobrevivência semelhantes e a presença de um platô próximo. Já a Figura 43 apresenta o gráfico do EKM estratificado por categoria de nódulo, exibindo curvas de sobrevivência distintas e um platô que se reduz progressivamente à medida que a categoria aumenta. Por fim, a Figura 44 apresenta o EKM estratificado por status de performance, evidenciando curvas distintas durante a maior parte do tempo. Observa-se, contudo, que ambas convergem para um platô semelhante, embora esse comportamento se inicie mais precocemente no grupo outro.

Assim como na aplicação anterior foram considerados na modelagem dos dados as distribuições de incidência Bernoulli ($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Ber}(\theta_{ijk})$), Poisson ($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Poisson}(\theta_{ijk})$) e geométrica ($M_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_{ijk})$), as distribuições da latência log-normal ($\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim} \text{LN}(\eta, \phi)$) e log-t-Student ($\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim} \text{LT}(\eta, \phi)$) com graus de liberdade 3, 4, 5 e 6 e os esquemas de pri-

meira ativação ($R_{ijk} = 1$), ativação aleatória ($R_{ijk} \stackrel{\text{ind}}{\sim} U\{1 \dots M_{ijk}\}$) e ativação geométrica ($R_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\xi)$) para o modelo em que T_{ijk} o tempo em anos de sobrevivência sem recidiva para o i -ésimo indivíduo com j -ésimo tratamento e k -ésimo sexo e supondo um modelo com esquema de ativação R_{ijk} , incidência M_{ijk} , latência $\tilde{Y}_{ijk}$ e com parte sistemática dada em θ_{ijk} , cuja **FSP** é dada por

$$S_{pop}(t_{ijk}) = P_{\theta_{ijk}}(R_{ijk} > M_{ijk}) \\ + E_{R_{ijk}, M_{ijk}} \left[IB(S_{\tilde{Y}_{ijk}}; M_{ijk} - R_{ijk} + 1, R_{ijk}) I(M_{ijk} \geq R_{ijk} \geq 1) \right]$$

e

$$b^{-1}(\theta_{ijk}) = \beta_0 + \beta_1 \text{idade}_{ijklm} + \beta_2 \text{breslow}_{ijklm} + \gamma_j + \alpha_k + \lambda_l + \tau_m, \quad j, k, l = 1, 2 \\ m = 1, 2, 3, 4, i = 1 \dots 426,$$

em que β_0 é o intercepto e β_1 e β_2 são os efeitos da idade_{ijklm} e da espessura_{ijklm} do tumor, com idade_{ijklm} e espessura_{ijklm} denotando a idade e a espessura do tumor do i -ésimo indivíduo com j -ésimo tratamento, k -ésimo sexo, l -ésima performance e m -ésima categoria de nódulo, respectivamente. Além disso γ_j é o efeito do j -ésimo tratamento, α_k é o efeito do k -ésimo sexo, λ_l é o efeito da l -ésima performance e τ_m é o efeito da m -ésima categoria de nódulo. Assume-se também que $\gamma_1 = 0$, $\alpha_1 = 0$, $\lambda_1 = 0$ e $\tau_1 = 0$ como casela de referência. A estimação dos parâmetros também foi feita usando o método de estimação por máxima verossimilhança utilizando a verossimilhança marginal e a escolha do modelo pelos critérios de menor **AIC** e **BIC**, assim como a aplicação anterior. Os **TRV** para os modelos com diferentes estruturas de incidência e latência resulta na hipótese $H_0: \xi = 1$ sendo rejeitada ao nível de significância de 5% somente nos casos com incidência Bernoulli. Assim, foram considerados os modelos com primeira ativação, ativação aleatória e ativação geométrica apenas com incidência Bernoulli.

Tabela 26 – Modelos ajustados com diferentes esquemas de ativação

Esquema	Incidência (M)	Latência (Y)	AIC	BIC
Primeira	Geométrica	log-t-Student ($\nu = 3$)	667,3882	711,7522
Primeira	Poisson	log-t-Student ($\nu = 3$)	670,4265	714,7905
Primeira	Geométrica	log-t-Student ($\nu = 4$)	671,1343	715,4982
Primeira	Geométrica	log-t-Student ($\nu = 5$)	674,1031	718,4670
Aleatória	Poisson	log-t-Student ($\nu = 3$)	674,8423	719,2063
Primeira	Poisson	log-t-Student ($\nu = 4$)	675,0497	719,4137
Primeira	Geométrica	log-t-Student ($\nu = 6$)	676,4462	720,8101
Primeira	Poisson	log-t-Student ($\nu = 5$)	678,3466	722,7106
Primeira	Poisson	log-t-Student ($\nu = 6$)	680,8108	725,1748
Aleatória	Poisson	log-t-Student ($\nu = 5$)	692,7359	737,0998
Primeira	Geométrica	log-normal	693,8686	738,2326
Aleatória	Poisson	log-t-Student ($\nu = 6$)	694,5895	738,9535
Primeira	Poisson	log-normal	697,6369	742,0008
Geométrica	Bernoulli	log-t-Student ($\nu = 3$)	706,4079	754,8049
Geométrica	Bernoulli	log-t-Student ($\nu = 4$)	711,6911	760,0881
Geométrica	Bernoulli	log-t-Student ($\nu = 5$)	715,2599	763,6569
Geométrica	Bernoulli	log-t-Student ($\nu = 6$)	717,8413	766,2384
Aleatória	Poisson	log-t-Student ($\nu = 4$)	730,5738	774,9378
Geométrica	Bernoulli	log-normal	734,2534	782,6504
Primeira	Bernoulli	log-t-Student ($\nu = 6$)	750,7730	795,1370
Aleatória	Geométrica	log-t-Student ($\nu = 6$)	750,7730	795,1370
Primeira	Bernoulli	log-t-Student ($\nu = 5$)	751,0287	795,3927
Aleatória	Geométrica	log-t-Student ($\nu = 5$)	751,0287	795,3927
Aleatória	Geométrica	log-t-Student ($\nu = 4$)	751,5860	795,9500
Primeira	Bernoulli	log-t-Student ($\nu = 4$)	751,5860	795,9500
Aleatória	Geométrica	log-normal	751,7252	796,0892
Primeira	Bernoulli	log-normal	751,7252	796,0892
Aleatória	Poisson	log-normal	751,7252	796,0892
Primeira	Bernoulli	log-t-Student ($\nu = 3$)	752,8665	797,2305
Aleatória	Geométrica	log-t-Student ($\nu = 3$)	752,8665	797,2305

Fonte: o autor

De acordo com a Tabela 26, o modelo com melhor desempenho segundo ambas as métricas é o de primeira ativação, incidência geométrica e latência log-t-Student com 3 graus de liberdade, sendo, portanto, o escolhido.

Assim, o modelo de ativação $R_{ijk} = 1$, incidência $M_{ijk} \stackrel{\text{ind}}{\sim} \text{Geom}(\theta_{ijk})$ e latência $\tilde{Y}_{ijk} \stackrel{\text{i.i.d.}}{\sim}$

$LT(3)$ com parte sistemática dada em θ_{ijk} tem sua **FSP** dada por

$$S_{pop}(t) = \frac{\theta_{ijklm}}{1 - (1 - \theta_{ijklm})S_Y(t)},$$

e

$$\log\left(\frac{\theta_{ijklm}}{1 - \theta_{ijklm}}\right) = \beta_0 + \beta_1 \text{idade}_{ijklm} + \beta_2 \text{breslow}_{ijklm} + \gamma_j + \alpha_k + \lambda_l + \tau_m, \quad j, k, l = 1, 2, \\ m = 1, 2, 3, 4, i = 1 \dots 426,$$

Nota-se que o modelo selecionado possui a mesma distribuição para as componentes de incidência e latência da aplicação anterior, o que é coerente, uma vez que os indivíduos da amostra apresentam as mesmas características clínicas e a mesma doença dos indivíduos da análise anterior. Por outro lado, a distribuição associada ao esquema de ativação, que nesse contexto pode ser interpretada como o nível de defesa do sistema imunológico, foi alterada. Anteriormente, adotava-se o esquema de ativação aleatória, no qual a ocorrência de morte ou recidiva tinha igual probabilidade de ocorrer por qualquer uma das causas da incidência M . No modelo atual, adotou-se o esquema de primeira ativação, em que a ocorrência de morte ou recidiva está associada à primeira causa que se manifesta. Essa mudança no mecanismo de ativação pode, ou não, estar relacionada à alteração da dosagem do tratamento.

Tabela 27 – Estimativas dos coeficientes do modelo em investigação

Coeficiente	Estimativa	EP	Z	Valor-p	IC 2,5%	IC 97,5%
η	2,3971	0,1935	12,3885	< 0,001	2,0179	2,7764
ϕ	0,3168	0,0526	6,0246	< 0,001	0,2137	0,4199
β_0 (Intercepto)	1,4277	0,5061	2,8211	0,005	0,4358	2,4195
γ_1 (tratamento)	-0,0024	0,1942	-0,0122	0,990	-0,3830	0,3783
τ_1 (nódulo = 2)	-0,6159	0,2893	-2,1291	0,033	-1,1829	-0,0489
τ_2 (nódulo = 3)	-0,9569	0,3018	-3,1701	0,002	-1,5485	-0,3653
τ_3 (nódulo = 4)	-1,6648	0,2993	-5,5625	< 0,001	-2,2515	-1,0782
α_1 (sexo masc.)	0,1852	0,2020	0,9167	0,359	-0,2108	0,5812
λ_1 (perform. ativo)	-0,1865	0,2825	-0,6602	0,509	-0,7401	0,3672
β_1 (idade)	-0,0132	0,0075	-1,7612	0,078	-0,0279	0,0015
β_2 (Breslow)	-0,0325	0,0315	-1,0315	0,302	-0,0943	0,0293

Fonte: o autor

A Tabela **27** mostra os resultados do ajuste do modelo com todas as covariáveis. Observou-se que apenas a covariável nódulos é significativa ao nível de 5%. O **AIC**, segundo o procedimento *stepwise*, indica que o modelo contendo nódulos e idade é o mais adequado, enquanto o **BIC** favorece o modelo com apenas nódulos.

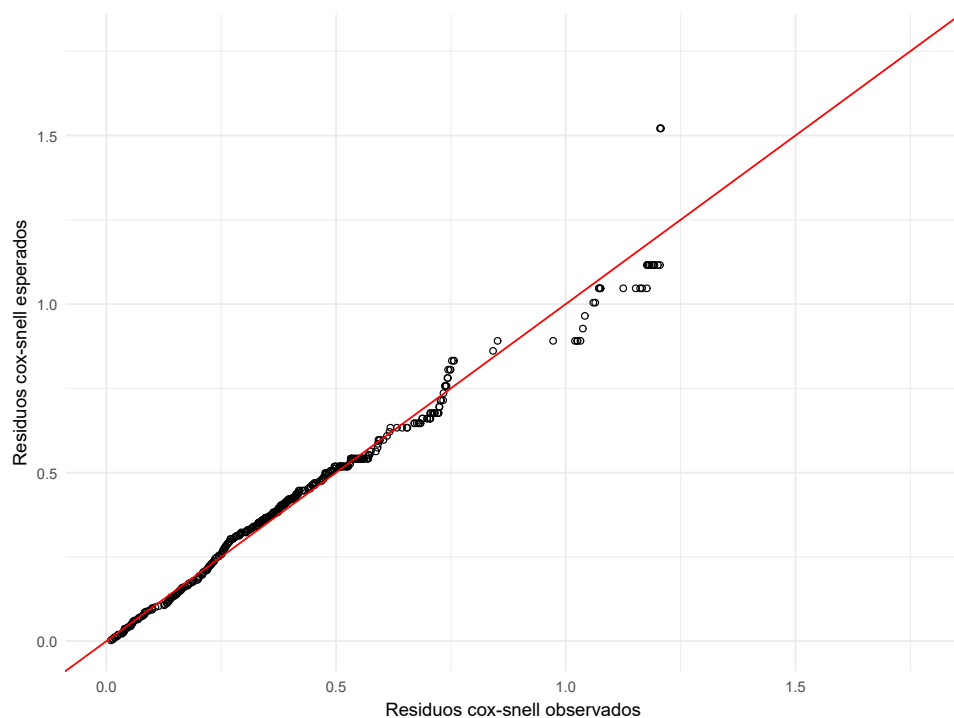
Tabela 28 – Estimativas dos coeficientes do modelo

Coeficiente	Estimativa	EP	Z	Valor-p	IC 2,5%	IC 97,5%
η	2,3712	0,1890	12,5482	< 0,001	2,0008	2,7416
ϕ	0,3191	0,0530	6,0246	< 0,001	0,2153	0,4230
β_0 (Intercepto)	0,5806	0,2066	2,8109	0,005	0,1758	0,9855
τ_1 (nódulo = 2)	-0,4025	0,2595	-1,5511	0,121	-0,9111	0,1061
τ_2 (nódulo = 3)	-0,7907	0,2843	-2,7811	0,005	-1,3480	-0,2335
τ_3 (nódulo = 4)	-1,5356	0,2802	-5,4803	< 0,001	-2,0848	-0,9864

Fonte: o autor

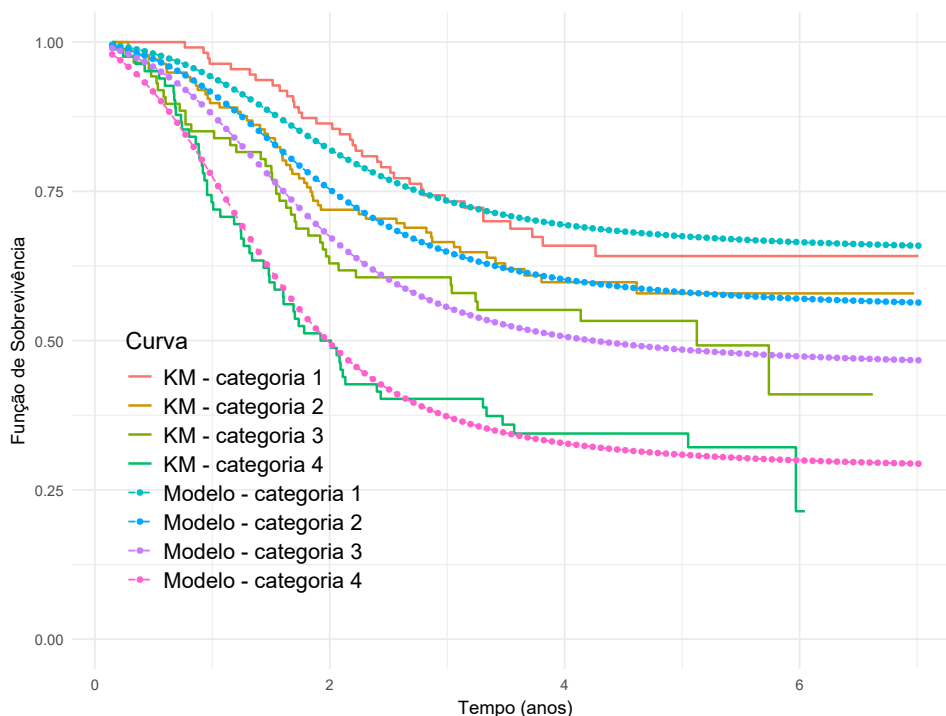
Realizando o **TRV** ao comparar o modelo com apenas a covariável nódulos com aquele que inclui nódulos e idade, o valor- p é 0,0676, sugerindo que não há evidência suficiente para rejeitar, ao nível de 5% de significância o modelo mais simples apenas com a covariável nódulos. Assim, o modelo final selecionado é aquele com a covariável nódulos e o intercepto. Os resultados desse modelo estão apresentados na Tabela **28** onde a categoria 2 da covariável nódulos não é significativa ao nível de 5% de significância, mas as 3 categorias conjuntamente apresentaram valor- p < 0,001, assim a covariável permaneceu no modelo.

Figura 45 – Gráfico dos resíduos de Cox-Snell.



Fonte: o autor

Figura 46 – Gráfico Kaplan-Meier por categoria de nódulos.



Fonte: o autor

A Figura 45 mostra os resíduos de Cox–Snell observados versus os esperados. Verifica-se, novamente, que o modelo está adequadamente especificado, com a maior parte dos pontos sobre ou próximos à linha de referência e apenas uma observação consideravelmente afastada, o que pode indicar um ponto discrepante.

Já a Figura 46 compara as curvas do EKM para as quatro categorias de nódulo com aquelas estimadas pelo modelo. De modo geral, o ajuste é considerado bom, pois as curvas estimadas pelo modelo estão próximas das curvas estimadas pelo EKM, especialmente nos últimos anos para as categorias 1 e 2 indicando uma boa modelagem da fração de curados nessas categorias.

Dessa forma, o modelo final ajustado para esses dados é da forma:

$$S_{pop}(t_{im}) = \frac{\theta_m}{1 - (1 - \theta_m)S_Y(t_{im})}$$

e

$$\log\left(\frac{\theta_m}{1 - \theta_m}\right) = \beta_0 + \tau_m, \quad m = 1, 2, 3, 4, \quad i = 1 \dots 426.$$

Esse modelo apresenta probabilidades de cura decrescentes conforme a categoria do nódulo

piora, conforme:

$$\hat{P}(\text{"Curado"} \mid \text{Categoria do n\u00f3dulo} = 1) = \hat{\theta}_1 = 0,6412,$$

$$\hat{P}(\text{"Curado"} \mid \text{Categoria do n\u00f3dulo} = 2) = \hat{\theta}_2 = 0,5444,$$

$$\hat{P}(\text{"Curado"} \mid \text{Categoria do n\u00f3dulo} = 3) = \hat{\theta}_3 = 0,4477,$$

$$\hat{P}(\text{"Curado"} \mid \text{Categoria do n\u00f3dulo} = 4) = \hat{\theta}_4 = 0,2779.$$

Observa-se que as probabilidades de cura diminuem de forma consistente, cerca de 10%, conforme a categoria do n\u00f3dulo se agrava. A \u00fanica exce\u00e7\u00e3o ocorre na categoria mais severa, na qual h\u00e1 uma queda mais acentuada, de aproximadamente 17%. Logo, quanto maior a quantidade de n\u00f3dulos, menor a probabilidades de cura. A raz\u00e3o de chances quantifica essa redu\u00e7\u00e3o: comparando-se com um paciente sem n\u00f3dulo, temos que, para um paciente com 1 n\u00f3dulo, $\hat{\psi}_{\tau_2} = 0,6686$, a chance de cura \u00e9 cerca de 33,14% menor. Para um paciente com 2 ou 3 n\u00f3dulos, com $\hat{\psi}_{\tau_3} = 0,4535$, a redu\u00e7\u00e3o \u00e9 de aproximadamente 54,65%, e para um paciente com mais que 3 n\u00f3dulos, com $\hat{\psi}_{\tau_4} = 0,2153$, a queda atinge cerca de 78,47%. Por fim, os intervalos de confian\u00e7a de 95% para $\hat{\psi}_{\tau_2}$, $\hat{\psi}_{\tau_3}$ e $\hat{\psi}_{\tau_4}$ s\u00e3o, respectivamente, de 0,4021 a 1,1121, de 0,2598 a 0,7918 e de 0,1242 a 0,3733. Assim, apenas para a categoria 2 n\u00e3o h\u00e1 evid\u00eancias suficientes para afirmar que a categoria exerce influ\u00eancia sobre o tempo at\u00e9 a recidiva.

6 CONSIDERAÇÕES FINAIS

Nesta dissertação generalizamos o modelo de fração de cura log-simétrico de Rocha, Me-deiros e Valença (2022) para diferentes esquemas de ativação. Foram derivadas as expressões da verossimilhança completa, bem como suas primeiras e segundas derivadas, além da verossimilhança marginal, para os casos particulares de esquemas de primeira ativação, ativação aleatória e ativação geométrica, considerando distribuições de incidência Bernoulli, Poisson e Geométrica.

Avaliamos o desempenho dos estimadores de máxima verossimilhança para os esquemas de ativação mencionados, assumindo incidência Poisson e latência log-normal e log-t-Student com 3 graus de liberdade. As covariáveis foram incorporadas apenas na fração de cura, por meio do parâmetro da distribuição de incidência. Dois cenários foram analisados: um com uma covariável binária representando, por exemplo, grupo de tratamento e controle e outro com uma covariável contínua representando, por exemplo, variação do peso corporal após certo tratamento. Os resultados mostraram que, de modo geral, o aumento do tamanho amostral reduz o viés e a variabilidade dos estimadores, exceto para o estimador do parâmetro ξ no esquema de ativação geométrica.

Verificou-se também que o aumento da fração de cura acentua o impacto do crescimento da censura nos não curados sobre as métricas consideradas. Além disso, o estimador da mediana $\hat{\eta}$ apresentou alto viés e variabilidade para os esquemas de primeira ativação e ativação geométrica, com melhor desempenho sob a distribuição log-t-Student com 3 graus de liberdade. Com relação às covariáveis, os estimadores dos parâmetros de regressão β no esquema de ativação aleatória apresentaram viés e variabilidade elevadas no cenário 2, o que não aconteceu no cenário 1.

Um segundo estudo de simulação foi realizado para avaliar o desempenho dos resíduos de Cox–Snell global sob especificações incorretas na latência, na incidência e no esquema de ativação. Os resultados indicam que o resíduo possui desempenho variável na detecção de erros de especificação do modelo: em alguns cenários é eficaz, mas em outros apresenta baixa sensibilidade. Esquemas de primeira ativação e ativação geométrica geraram resultados semelhantes, dificultando a distinção entre eles mesmo sob especificação incorreta do esquema, ou seja, quando o modelo foi gerado por um esquema e estimado por outro. Além disso, no esquema de ativação aleatória, observaram-se desvios mesmo sob especificação correta, os

quais se agravaram em presença de erro na latência, incidência ou no esquema.

Por fim, foram analisados dados de dois estudos clínicos sobre melanoma cutâneo de alto risco. A modelagem desenvolvida nesta dissertação foi aplicada e os resultados confirmaram as conclusões dos artigos originais, ampliando a análise ao permitir modelar a probabilidade de cura dos indivíduos em função das covariáveis sob diferentes esquemas de ativação.

REFERÊNCIAS

- ASSUNÇÃO, D. M. X. *Modelos de fração de cura com fragilidade inflacionado de zero sob diferentes esquemas de ativação*. Tese (Doutorado em Estatística) — Universidade de São Paulo, São Paulo, 2023. Orientadora: Vera Lúcia Damasceno Tomazella. Corientador: Jeremias da Silva Leão. Disponível em: <https://www.teses.usp.br/teses/disponiveis/104/104131/tde-11092023-145419/pt-br.php>.
- BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, Taylor & Francis, v. 47, n. 259, p. 501–515, 1952. Disponível em: <https://www.jstor.org/stable/2281318>.
- BOAG, J. W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, v. 11, n. 1, p. 15–53, 1949. Disponível em: <https://www.jstor.org/stable/2983694>.
- CARNEIRO, H. P. d. A. *Testes de hipóteses em modelos de sobrevivência com fração de cura*. Dissertação (Mestrado em Estatística) — Universidade Federal do Rio Grande do Norte, Natal, 2012. Orientadora: Dione Maria Valença. Corientador: Silvia L. P. Ferrari. Disponível em: <https://repositorio.ufrn.br/items/f3ea3e2d-26d6-4778-8c8b-397ef6bad6a4>.
- CHEN, M.-H.; IBRAHIM, J. G.; SINHA, D. A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, Taylor & Francis, v. 94, n. 447, p. 909–919, 1999. Disponível em: <https://www.jstor.org/stable/2670006>.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Editora Blucher, 2006. Disponível em: https://www.blucher.com.br/analise-de-sobrevivencia-aplicada_9788521203841.
- COONER, F.; BANERJEE, S.; CARLIN, B. P.; SINHA, D. Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, Taylor & Francis, v. 102, n. 478, p. 560–572, 2007. Disponível em: <https://www.tandfonline.com/doi/abs/10.1198/016214507000000112>.
- CORBIERE, F.; JOLY, P. A sas macro for parametric and semiparametric mixture cure models. *Computer methods and programs in biomedicine*, Elsevier, v. 85, n. 2, p. 173–180, 2007. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169260706002513>.
- COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. Disponível em: <https://www.jstor.org/stable/2985181>.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, Oxford University Press, v. 30, n. 2, p. 248–275, 1968. Disponível em: <https://www.jstor.org/stable/2984505>.
- CUNHA, D. R.; DIVINO, J. A.; SAULO, H. On a log-symmetric quantile tobit model applied to female labor supply data. *Journal of Applied Statistics*, Taylor & Francis, v. 49, n. 16, p. 4225–4253, 2022. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/02664763.2021.1976120>.

CUNHA, D. R.; DIVINO, J. A.; SAULO, H. Zero-adjusted log-symmetric quantile regression models. *Computational Economics*, Springer, v. 63, n. 5, p. 2087–2111, 2024. Disponível em: <https://link.springer.com/article/10.1007/s10614-023-10420-4>.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977. Disponível em: <https://www.jstor.org/stable/2984875>.

FANG, K. W. *Symmetric multivariate and related distributions*. Chapman and Hall/CRC, 1990. Disponível em: <https://www.taylorfrancis.com/books/mono/10.1201/9781351077040/symmetric-multivariate-related-distributions-kai-tai-fang-samuel-kotz-kai-ng>.

FIDELIS, C. R.; ORTEGA, E. M.; CORDEIRO, G. M. Residual analysis for poisson-exponentiated weibull regression models with cure fraction. *Stats*, MDPI, v. 7, n. 2, p. 492–507, 2024. Disponível em: <https://www.mdpi.com/2571-905X/7/2/30>.

JOBÉ, J. M. *Lognormal distributions: Theory and applications*. New York: Taylor & Francis, 1989. Disponível em: <https://academic.oup.com/jrsssc/article/38/2/381/6985570>.

JONES, M. On reciprocal symmetry. *Journal of Statistical Planning and Inference*, Elsevier, v. 138, n. 10, p. 3039–3043, 2008. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0378375807004053>.

JONES, M.; ARNOLD, B. C. Distributions that are both log-symmetric and r-symmetric. *Electronic Journal of Statistics*, Institute of Mathematical Statistics (IMS) and the Bernoulli Society., v. 2, p. 1300–1308, 2008. Disponível em: <https://arxiv.org/abs/0810.0102>.

KALBFLEISCH, J. D.; PRENTICE, R. L. *The statistical analysis of failure time data*. New Jersey: John Wiley & Sons, 2011. Disponível em: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118032985>.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Disponível em: <https://www.jstor.org/stable/2281868>.

KIM, S.; CHEN, M.-H.; DEY, D. K. A new threshold regression model for survival data with a cure fraction. *Lifetime data analysis*, Springer, v. 17, p. 101–122, 2011. Disponível em: <https://link.springer.com/article/10.1007/s10985-010-9166-9>.

KIRKWOOD, J. M.; IBRAHIM, J. G.; SONDAK, V. K.; RICHARDS, J.; FLAHERTY, L. E.; ERNSTOFF, M. S.; SMITH, T. J.; RAO, U.; STEELE, M.; BLUM, R. H. High-and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial e1690/s9111/c9190. *Journal of clinical oncology*, American Society of Clinical Oncology, v. 18, n. 12, p. 2444–2458, 2000. Disponível em: https://ascopubs.org/doi/10.1200/JCO.2000.18.12.2444?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed.

KIRKWOOD, J. M.; STRAWDERMAN, M. H.; ERNSTOFF, M. S.; SMITH, T. J.; BORDEN, E. C.; BLUM, R. H. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial est 1684. *Journal of clinical oncology*, American Society of Clinical Oncolog, v. 14, n. 1, p. 7–17, 1996. Disponível em: https://ascopubs.org/doi/10.1200/JCO.1996.14.1.7?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed.

KLEINBAUM, D. G.; KLEIN, M. *Survival analysis a self-learning text*. New York: Springer, 1996. Disponível em: <https://link.springer.com/book/10.1007/978-1-4419-6646-9>.

LI, C.-S.; TAYLOR, J. M. A semi-parametric accelerated failure time cure model. *Statistics in medicine*, Wiley Online Library, v. 21, n. 21, p. 3235–3247, 2002. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1260>.

MALLER, R. A.; ZHOU, X. *Survival analysis with long-term survivors*. New York: John Wiley Sons, 1996. Disponível em: <https://www.wiley.com/en-us/Survival+Analysis+with+Long-Term+Survivors-p-9780471962014>.

MCCULLAGH, P.; NELDER, J. *Generalized Linear Models*. Springer, 1983. (Monographs on Statistics and Applied Probability). Disponível em: <https://books.google.com.br/books?id=OUitAQAACAAJ>.

MEEKER, W. Q. Limited failure population life tests: application to integrated circuit reliability. *Technometrics*, Taylor & Francis, v. 29, n. 1, p. 51–65, 1987. Disponível em: <https://www.jstor.org/stable/1269883>.

NELSON, W. B. *Accelerated testing: statistical models, test plans, and data analysis*. Malden: John Wiley & Sons, 2009. Disponível em: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316795>.

PAIVA, L. d. S. *Uma classe de distribuições log-simétricas discretas*. Dissertação (Mestrado em Estatística) — Universidade de Brasília, Brasília, 2020. Orientadora: Helton Saulo Bezerra dos Santos. Corientador: Roberto Vila gabriel. Disponível em: <https://repositorio.unb.br/handle/10482/38584?locale=en>.

PENG, Y.; TAYLOR, J. M. Residual-based model diagnosis methods for mixture cure models. *Biometrics*, Wiley Online Library, v. 73, n. 2, p. 495–505, 2017. Disponível em: <https://www.jstor.org/stable/44695174>.

ROCHA, J. B.; MEDEIROS, F. M.; VALENÇA, D. M. Log-symmetric models with cure fraction with application to leprosy reactions data. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 36, n. 3, p. 560–578, 2022. Disponível em: <https://projecteuclid.org/journals/brazilian-journal-of-probability-and-statistics/volume-36/issue-3/Log-symmetric-models-with-cure-fraction-with-application-to-leprosy/10.1214/22-BJPS540.short>.

RODRIGUES, J.; CANCHO, V. G.; CASTRO, M. de; LOUZADA-NETO, F. On the unification of long-term survival models. *Statistics & Probability Letters*, Elsevier, v. 79, n. 6, p. 753–759, 2009. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167715208005105>.

SAULO, H.; DASILVA, A.; LEIVA, V.; SÁNCHEZ, L.; FUENTE-MELLA, H. de la. Log-symmetric quantile regression models. *Statistica Neerlandica*, Wiley Online Library, v. 76, n. 2, p. 124–163, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/stan.12243>.

SCHMIDT, P.; WITTE, A. D. Predicting criminal recidivism using ‘split population’ survival time models. *Journal of Econometrics*, Elsevier, v. 40, n. 1, p. 141–159, 1989. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0304407689900341>.

SCOLAS, S.; LEGRAND, C.; OULHAJ, A.; GHOUCH, A. E. Diagnostic checks in mixture cure models with interval-censoring. *Statistical Methods in Medical Research*, SAGE Publications Sage UK: London, England, v. 27, n. 7, p. 2114–2131, 2018. Disponível em: https://journals.sagepub.com/doi/10.1177/0962280216676502?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed.

SESHADRI, V. On random variables which have the same distribution as their reciprocals. *Canadian Mathematical Bulletin*, Cambridge University Press, v. 8, n. 6, p. 819–824, 1965. Disponível em: <https://www.cambridge.org/core/journals/canadian-mathematical-bulletin/article/on-random-variables-which-have-the-same-distribution-as-their-reciprocals/1AABB6D2C79340C56D96B94140A892F4>.

TEAM, R. R. C. et al. *R: A language and environment for statistical computing*. Citeseer, 2020. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bc748aec2bcc46eb319d5446da614b37729e9dbe>.

THERNEAU, T. M. *Modeling Survival Data: Extending the cox model*. Malden: Springer, 2000. Disponível em: <https://link.springer.com/book/10.1007/978-1-4757-3294-8>.

THERNEAU, T. M.; GRAMBSCH, P. M.; FLEMING, T. R. Martingale-based residuals for survival models. *Biometrika*, Oxford University Press, v. 77, n. 1, p. 147–160, 1990. Disponível em: <https://www.jstor.org/stable/2336057>.

TSODIKOV, A. D.; YAKOVLEV, A. Y.; ASSELAIN, B. *Stochastic models of tumor latency and their biostatistical applications*. New Jersey: World Scientific, 1996. Disponível em: <https://www.worldscientific.com/worldscibooks/10.1142/2420?srsid=AfmBOoquJQiTkw-1Cbbwg-cq7iPVczuR3ULcbQ74IVGzHy7vjXvjaw2N#t=aboutBook>.

VANEGAS, L. H.; PAULA, G. A. Log-symmetric distributions: statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 30, n. 2, p. 196–220, 2016. Disponível em: <https://projecteuclid.org/journals/brazilian-journal-of-probability-and-statistics/volume-30/issue-2/Log-symmetric-distributions-Statistical-properties-and-parameter-estimation/10.1214/14-BJPS272.full>.

VILA, R.; BALAKRISHNAN, N.; SAULO, H.; PROTAZIO, A. Bivariate log-symmetric models: Distributional properties, parameter estimation and an application to public spending data. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 37, n. 3, p. 619–642, 2023. Disponível em: <https://projecteuclid.org/journals/brazilian-journal-of-probability-and-statistics/volume-37/issue-3/Bivariate-log-symmetric-models--Distributional-properties-parameter-estimation-and/10.1214/23-BJPS584.short>.

WILEYTO, E. P.; LI, Y.; CHEN, J.; HEITJAN, D. F. Assessing the fit of parametric cure models. *Biostatistics*, Oxford University Press, v. 14, n. 2, p. 340–350, 2013. Disponível em: <https://academic.oup.com/biostatistics/article-abstract/14/2/340/377666?redirectedFrom=fulltext>.

YIN, G.; IBRAHIM, J. G. Cure rate models: a unified approach. *Canadian Journal of Statistics*, Wiley Online Library, v. 33, n. 4, p. 559–570, 2005. Disponível em: <https://www.jstor.org/stable/25046202>.