



Relatório

Modelagem da persistência de produtos de seguro ofertados
pela MONGERAL AEGON via modelos de sobrevivência com
mistura

29 de abril de 2025



Conteúdo

1	Introdução	4
2	Análise de Sobrevivência	6
3	Modelo Bayesiano de mistura Log-Normal com presença de dados censurados	8
3.1	Métodos para agrupamentos	9
4	Análise do desfecho Resgate	11
4.1	Base de dados	11
4.1.1	Descrição dos dados	11
4.1.2	Remoções	14
4.1.3	Proporções de censura e resgate	15
4.1.4	Proporções por categoria do cliente	16
4.1.5	Proporções por forma de cobrança	17
4.1.6	Proporções por tipo de cobertura	18
4.1.7	Clusterização	19
4.2	Aplicação do Modelo de Mistura	21
4.2.1	Cenários Avaliados	21
4.2.2	Representatividade dos Cenários na Base	22
4.2.3	Parâmetros Estimados	23
4.2.4	Curvas de Persistência e Taxas de Falha	25
4.2.5	Comparações das Curvas de Persistência e Taxas de Falha para os cenários	30
4.3	Probabilidades de Persistência	33
4.3.1	Exemplo 1	33
4.3.2	Exemplo 2	34
4.3.3	Exemplo 3	34
5	Análise do desfecho Saldamento	36
5.1	Base de dados	36
5.1.1	Descrição dos dados	36
5.1.2	Remoções	38



5.1.3	Proporções de censura e saldamento	39
5.1.4	Proporções por categoria do cliente	40
5.1.5	Proporções por forma de cobrança	41
5.1.6	Proporções por tipo de cobertura	42
5.1.7	Clusterização	43
5.2	Aplicação do Modelo de Mistura	46
5.2.1	Cenários de Interesse Estudados	46
5.2.2	Representação dos Cenários nas Bases	47
5.2.3	Parâmetros Estimados	48
5.2.4	Curvas de Persistência e Taxas de Falha	50
5.2.5	Comparações das Curvas de Persistência e Taxas de Falha para os cenários	54
5.3	Probabilidades de Persistência	58
5.3.1	Exemplo 1	58
5.3.2	Exemplo 2	58
5.3.3	Exemplo 3	59
6	Considerações Finais	60
7	Equipe Responsável	61
A	Descrição completa das covariáveis e dos clusters	62
A.1	Desfecho Resgate: Conjunto de covariáveis	62
A.2	Desfecho Resgate: Descrição dos clusters por covariável	62
A.3	Desfecho Saldamento: Conjunto de covariáveis	63
A.4	Desfecho Saldamento: Descrição dos clusters por covariável	63
B	Descrição da Censura	64



1 Introdução

Neste relatório, daremos continuidade ao estudo sobre o comportamento de persistência em produtos de seguros, analisando os fatores que influenciam as decisões dos segurados. Utilizaremos os mesmos conceitos e metodologias de análise de sobrevivência discutidos no relatório anterior, abordando como o tempo até a falha, em produtos de capitalização, pode ser influenciado por dois desfechos distintos: resgate, que se refere à retirada do valor acumulado pelo cliente, encerrando ou reduzindo o saldo do contrato, e saldamento, que ocorre quando o cliente interrompe o pagamento dos prêmios, mas mantém a cobertura com um valor reduzido proporcionalmente ao acumulado. Este aprofundamento é crucial para um melhor entendimento das dinâmicas do mercado e dos desafios que podem ser enfrentados pela seguradora.

Buscaremos expandir a análise utilizando a modelagem paramétrica com mistura de distribuições Log-Normal (MSMLN), conforme já previamente introduzido e proposto em [Lobo et al. \[2024\]](#). Este modelo é capaz de capturar mudanças bruscas das curvas de sobrevivência ao longo da janela de estudo, característica bastante marcante da base de dados analisada. A flexibilidade do MSMLN se destaca em comparação aos modelos usuais de sobrevivência já bem estabelecidos na literatura. Além disso, é importante ressaltar a presença de altas taxas de censura nos dados, já que grande parte dos produtos não é resgatada ou saldada até o fim do estudo. Ao final do período, a inscrição do cliente pode ainda estar ativa, resultando em taxas de censura que variam entre 40% e 95%. Essa alta taxa de censura representa um grande desafio, mas o modelo proposto é suficientemente flexível para abordar essa questão de forma eficaz e a abordagem computacional otimizada para tratar essa particularidade. A janela de estudo inicia-se em 2010 e estende-se até 2023, abrangendo um período total de 168 meses, que corresponde a 14 anos de dados analisados, permitindo uma avaliação robusta das tendências de sobrevivência.

Toda a análise foi realizada utilizando o pacote **lnmixsurv**, desenvolvido em linguagem R [[R Core Team, 2023](#)]. Este pacote, criado pela equipe do LabMA/UFRJ e disponível no CRAN-R, fornece a modelagem de modelos de mistura de sobrevivência Log-Normal através da abordagem Bayesiana, via algoritmo de Expectation-Maximization. Uma descrição detalhada das funcionalidades e aplicações do pacote pode ser encontrada em [Lobo et al. \[2024\]](#) e [GitHub | lnmixsurv](#).



O restante deste documento está organizado da seguinte forma: na Seção 2 apresentamos a análise de sobrevivência. Na Seção 3, discutimos o modelo Bayesiano de mistura com resposta Log-Normal para dados censurados. Na sequência, analisamos o desfecho resgate na Seção 4 e o desfecho saldamento na Seção 5. Em ambas as seções, é apresentada a base de dados, abrangendo a descrição dos dados, as proporções de censura, e o processo de clusterização. Além disso, na modelagem, incluiu-se cenários de interesse, bem como as curvas ajustadas de persistência e de risco. Além disso, considerou-se calcular as probabilidades de persistência para clientes em relação a ambos os desfechos, de forma a oferecer uma percepção sobre a probabilidade de um cliente continuar ativo na empresa ao longo do tempo, o que é essencial para estratégias de retenção e planejamento da empresa. Por fim, a Seção 6 conclui o documento, com as considerações finais.



2 Análise de Sobrevivência

Seja T uma variável aleatória contínua e não-negativa que denote o tempo de vida de indivíduos em uma determinada população de interesse. Seja $f(t)$ a função densidade de probabilidade de T e $F(t)$ a respectiva função de distribuição acumulada, representada pela seguinte expressão:

$$F(t) = P(T \leq t) = \int_0^t f(u) du. \quad (1)$$

A probabilidade de sobrevivência de um indivíduo até o tempo t , usualmente denotada por $S(t)$, é obtida pela função de sobrevivência:

$$S(t) = 1 - F(t) = P(T > t). \quad (2)$$

Note que a função $S(t)$ é monótona e decrescente, começando em $S(0) = 1$ e convergindo numericamente para zero, uma vez que $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$. Define-se agora o conceito de função taxa de falha, $h(t)$, que representa a taxa de falha instantânea em um certo tempo t , expressa como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad (3)$$

onde Δt é um pequeno acréscimo de tempo. Em particular, $h(t)\Delta t$ é a probabilidade aproximada de ocorrer uma falha no intervalo $(t, t + \Delta t)$, dado que houve sobrevivência até o tempo t . As funções $f(t)$, $F(t)$, $S(t)$ e $h(t)$ fornecem especificações matematicamente equivalentes a distribuição de T . Além disso, a partir de $S(t)$ e $f(t)$, é possível obter uma expressão para $h(t)$:

$$h(t) = -\frac{d}{dt} \log S(t) \quad (4)$$

Dentro do contexto de análise de sobrevivência, modelos paramétricos desempenham um papel fundamental na modelagem do fenômeno de interesse. As distribuições usuais deste contexto são: Exponencial, Weibull e Log-Normal. [Colosimo and Giolo \[2006\]](#)



Para o presente estudo, a variável aleatória T representa o tempo até cancelamento de um determinado contrato oferecido por uma empresa de seguros, podendo ser resgate ou saldamento. Em estudos prévios, avaliou-se que a distribuição Log-Normal melhor se ajustava aos dados observados, em comparação às distribuições Exponencial e Weibull. Por outro lado, também verificou-se que o ajuste utilizando o modelo Log-Normal poderia ser melhorado, e para isso propôs-se o uso de modelos de mistura, conforme será apresentado adiante.

Modelo Log-Normal

Seja $\mathbf{T} = T_1, T_2, \dots, T_n$ uma amostra aleatória cuja distribuição de um certo elemento $T_i \in \mathbf{T}$ segue a distribuição Log-Normal com parâmetro de locação μ_i e parâmetro de dispersão σ , isto é, em notação: $T_i \sim \mathcal{LN}(\mu_i, \sigma^2)$, então

$$f(t_i | \mu_i, \sigma^2) = (2\pi)^{-\frac{1}{2}} (t_i \sigma)^{-1} \exp \left\{ -\frac{1}{2\sigma^2} (\log(t_i) - \mu_i)^2 \right\}. \quad (5)$$

Como o documento será desenvolvido a partir de uma regressão linear para a variável aleatória T , então, naturalmente, é feito a inclusão de covariáveis contextuais que estão associadas ao fenômeno do cancelamento. Neste sentido, tem-se uma matriz de dados \mathbf{X} a partir do qual o preditor linear $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é construído. Consequentemente, a expressão em 5 pode ser reescrita como

$$f(t_i | \mu, \sigma^2) = (2\pi)^{-\frac{1}{2}} (t_i \sigma)^{-1} \exp \left\{ -\frac{1}{2\sigma^2} (\log(t_i) - \mathbf{x}_i^T \boldsymbol{\beta})^2 \right\}. \quad (6)$$

Por outro lado, a função de sobrevivência considerando o modelo Log-Normal com preditor linear $\mathbf{x}_i^T \boldsymbol{\beta}$ pode ser escrito como

$$\Phi \left(\frac{-\log(t_i) + \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right). \quad (7)$$

Tendo uma forma fechada da função de sobrevivência, apresentada na expressão 7, é possível obter a função taxa de falha $h(t)$ (equação 3 e 4) para a suposição Log-Normal no contexto de regressão linear.



3 Modelo Bayesiano de mistura Log-Normal com presença de dados censurados

O presente estudo foi desenvolvido com base no modelo de mistura Log-Normal proposto em [Lobo et al. \[2024\]](#), com objetivo principal de analisar o tempo até a falha de contratos de seguro, capturando a diversidade observada nos padrões de falha. O tempo até a ocorrência do evento segue uma distribuição Log-Normal, ou seja, $Y_i = \log(T_i)$, onde $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$. Além disso, cada contrato de seguro é associado a um indicador de grupo latente $I_i = g$, $g \in \{1, \dots, G\}$, indicando que a i -ésima observação pertence à componente de mistura g ([Frühwirth-Schnatter \[2006\]](#)). Os grupos são tratados como variáveis latentes, não sendo conhecidos *a priori*, o que permite a captura da heterogeneidade nos dados. Temos que cada componente de mistura g assume uma proporção de mistura η_g , onde

$$\eta_g \in [0, 1], \quad \sum_{g=1}^G \eta_g = 1. \quad (8)$$

A função de verossimilhança aumentada para o modelo de misturas é dada por

$$f(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}, \mathbf{I}) = \prod_{g=1}^G \eta_g^{n_g} \left[\prod_{i: I_i=g} \mathcal{N}_g(y_i \mid \mathbf{x}_i^T \boldsymbol{\beta}_g, \phi_g^{-1}) \right], \quad (9)$$

em que \mathbf{I} representa os grupos latentes, n_g o número de observações e \mathcal{N}_g a distribuição Normal, ambos para a componente g .

Definimos $\mathbf{y} = (\mathbf{y}^{obs}, \mathbf{z})$, sendo \mathbf{y}^{obs} os logaritmos dos tempos até o resgate ou o saldamento e \mathbf{z} os logaritmos dos tempos censurados. A metodologia de aumento de dados permite simular tempos de saldamento ou resgate para os dados censurados, tratando-os como ocorrências, resultando em um vetor \mathbf{y}^{aug} tal que

$$y_i^{aug} = \begin{cases} y_i^{obs}, & \text{se } \delta_i = 1; \\ \text{amostrar de } \mathcal{NT}_{(z_i, +\infty)}(\mathbf{x}_i^T \boldsymbol{\beta}_g, \phi_g^{-1}), & \text{se } \delta_i = 0. \end{cases} \quad (10)$$

onde $\mathcal{NT}_{(z_i, +\infty)}(\mathbf{x}_i^T \boldsymbol{\beta}_g, \phi_g^{-1})$ é a distribuição Normal truncada inferiormente no tempo de censura z_i , de média $\mathbf{x}_i^T \boldsymbol{\beta}_g$ e variância ϕ_g^{-1} ; \mathbf{y}^{aug} substitui \mathbf{y} na função de verossimilhança dada pela equação (9). Para mais detalhes sobre a metodologia de aumento de dados ver [Tanner and Wong \[1987\]](#).



São definidas distribuições *a priori* para os parâmetros do modelo, como os pesos da mistura e os parâmetros dos componentes. Essas distribuições refletem o conhecimento prévio sobre os parâmetros. As distribuições *a priori* são as seguintes:

$$y_i^{aug} \mid \mathbf{X}, I_i, \boldsymbol{\beta}_g, \phi_g \sim \mathcal{N}_g(\mathbf{x}_i^T \boldsymbol{\beta}_g, \phi_g^{-1}) \quad (11)$$

$$I_i \mid \boldsymbol{\eta} \sim \text{Categórica}(G, \boldsymbol{\eta}) \quad (12)$$

$$\boldsymbol{\eta} \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_G) \quad (13)$$

$$\boldsymbol{\beta}_g \sim \mathcal{N}_g(\mathbf{m}_g, \mathbf{I}_p \sigma_g^2) \quad (14)$$

$$\phi_g \sim \text{Gama}(a_g, b_g), \quad (15)$$

em que \mathbf{I}_p é uma matriz identidade de ordem p .

Inferência Bayesiana é conduzida para estimar os parâmetros do modelo e obter a distribuição *a posteriori*. Métodos como o algoritmo de Metropolis-Hastings e MCMC são empregados para explorar o espaço de parâmetros ([Gamerman and Lopes \[2006\]](#)).

Além das contribuições destacadas em [Lobo et al. \[2024\]](#), é importante mencionar que a pesquisa resultou no desenvolvimento do pacote **lnmixsurv**. Criado em linguagem R e otimizado em C++ para eficiência computacional, o pacote **lnmixsurv** não apenas facilita, mas também disponibiliza ao público a aplicação da metodologia proposta. Este pacote, disponível no CRAN, possui implementação do amostrador de Gibbs e utiliza o algoritmo Expectation-Maximization para aprimorar a busca por melhores valores iniciais. Ressalta-se que toda a modelagem apresentada neste relatório foi realizada pelo pacote **lnmixsurv**.

3.1 Métodos para agrupamentos

Nas seções a seguir poderemos observar que algumas covariáveis potenciais para a análise de persistência possuem muitas categorias. Sendo assim, podemos definir *clusters* de forma a identificar categorias semelhantes nas covariáveis, facilitando a análise dos produtos de capitalização. Utilizou-se o pacote **mclust** do software R, especificamente a função **mclustBIC**, que avalia diferentes configurações de modelos de mistura Gaussiana e o número de clusters, fornecendo o valor do Critério de Informação Bayesiano (BIC) para cada configuração.

O BIC é uma métrica utilizada para avaliar a qualidade do ajuste, evitando sobreajustes



e indicando a seleção de um modelo mais parcimonioso e eficiente. A partir dos valores de BIC, é possível identificar a configuração que melhor se adapta aos dados, garantindo uma modelagem robusta e precisa para a identificação dos *clusters*. Essa abordagem permite não só otimizar a segmentação dos dados, mas também criar agrupamentos que refletem com maior precisão as características dos produtos de capitalização em análise.



4 Análise do desfecho Resgate

4.1 Base de dados

Nesta subseção, será apresentada a descrição do conjunto de dados fornecido pela empresa, destacando os agrupamentos realizados nas categorias das covariáveis, bem como as remoções efetuadas na base de dados.

4.1.1 Descrição dos dados

A Tabela 1 apresenta um resumo das covariáveis utilizadas para descrever o processo de persistência, com a variável de interesse sendo o tempo (em meses) até o desfecho resgate. Todas as covariáveis são categóricas, tornando essencial entender o comportamento das categorias associadas a cada uma delas. Devido ao número elevado de categorias em cada covariável, propôs-se agrupar essas categorias por meio da clusterização, a fim de simplificar a análise, como descrito na subseção 3.1.

Tabela 1: Desfecho Resgate: Descrição das covariáveis.

Covariável	Tipo	Suporte
Categoria do Cliente	Categórica	$\{1, 2, 3\}$
Forma de Cobrança	Categórica	$\{1, \dots, 7\}$
Tipo de Cobertura	Categórica	$\{1, 2, 3\}$

As Figuras 1, 2 e 3 ilustram as funções de sobrevivência e risco associadas às covariáveis *Categoria do Cliente*, *Forma de Cobrança* e *Tipo de Cobertura*, respectivamente. Nessas figuras podemos obter indicativos das diferenças entre as categorias, facilitando a compreensão das covariáveis estudadas e justificando a necessidade de clusterização.

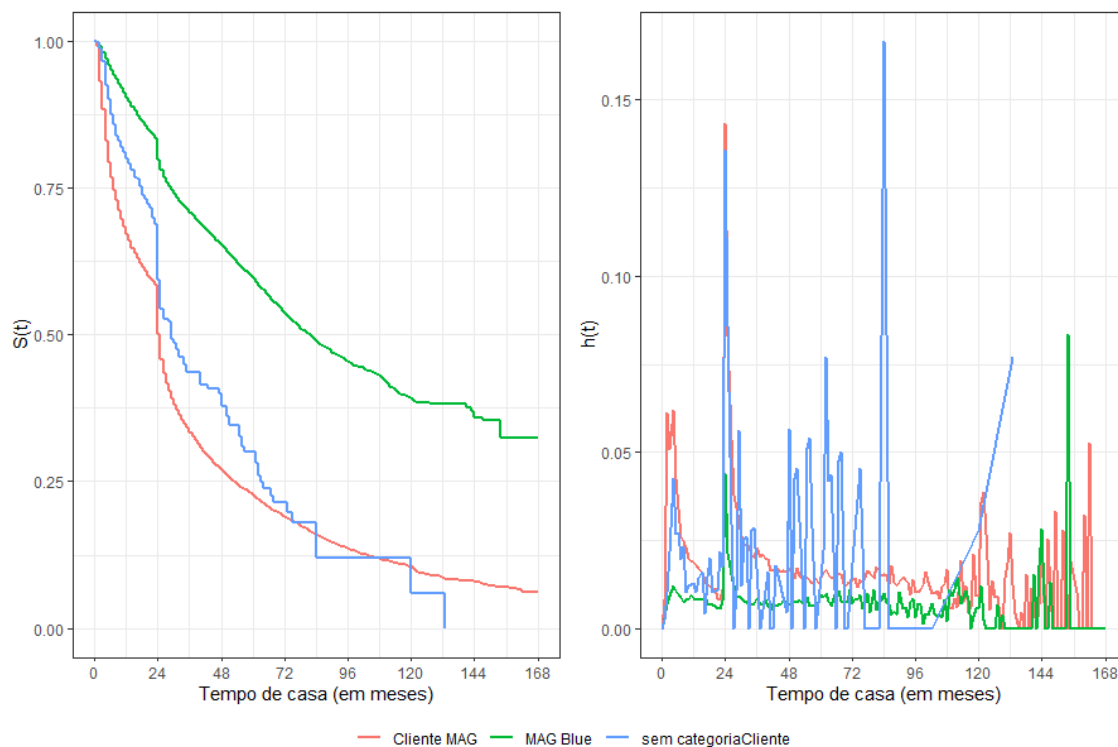


Figura 1: Desfecho Resgate: Função de Sobrevivência e Risco - Categoria do Cliente.

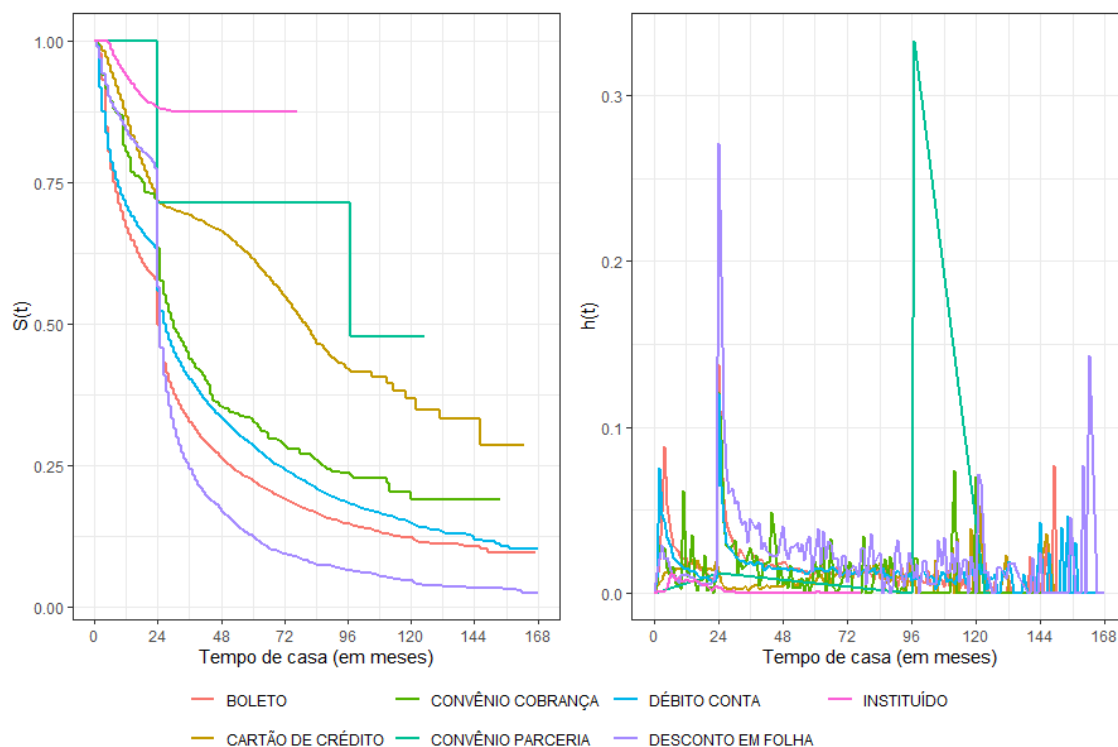


Figura 2: Desfecho Resgate: Função de Sobrevivência e Risco - Forma de Cobrança.

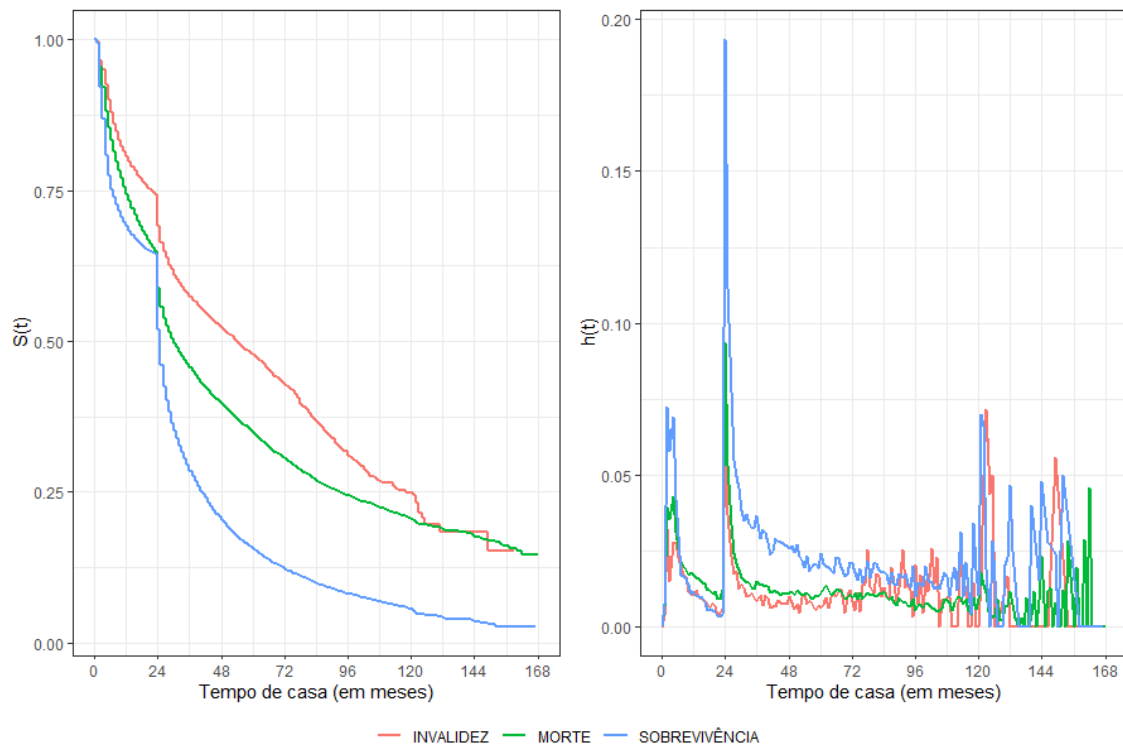


Figura 3: Desfecho Resgate: Função de Sobrevivência e Risco - Tipo de Cobertura.

Durante o estudo, foram analisadas outras covariáveis como *Sexo*, *Periodicidade de Cobrança* e *Cobertura Agrupamento*. No entanto, essas covariáveis não foram consideradas no estudo, pois a análise dos dados indicou que não apresentavam significância estatística relevante para o modelo.



4.1.2 Remoções

Antes de iniciar qualquer análise, foram realizadas remoções de alguns *status item contratado*. Parte dessas remoções foi efetuada com base em recomendações feitas pela própria MAG. Os valores absolutos e as proporções de cada *status item contratado* removidos estão detalhados na Tabela 2. Essas remoções foram aplicadas à base original, antes da separação entre as bases resgate e saldamento.

Tabela 2: Remoções dos status

Status Removidos	Valor Absoluto	Frequência (%)
CONTRATO ANULADO - C16	5.211	1,20
ENCERRAMENTO APÓLICE - C13	17	0
TRANSFERÊNCIA - C15	11.152	2,58

Além das remoções especificadas na Tabela 2, foram removidos itens com data de início da vigência da cobertura anterior a 2010. Foram removidos os *status item contratado*: *ORGÃO CONSIGNANTE CANCELADO*, *RENDA INVALIDEZ - B18*, *SOLICITAÇÃO RESGATE - C12*, *LIQUIDAÇÃO BENEFÍCIO - B04*, *PAGAMENTO ÚNICO - D01*, *PENSÃO - B01*, *REGULADO ICATU - C20* e *REMIDO VALOR INFERIOR MÍNIMO COBRANÇA - D03*, resultando na remoção de um total de 207.485 observações.

Outras remoções: Tempos até o desfecho

Antes de iniciar os estudos sobre persistência, foram realizadas algumas alterações na base de dados. Por recomendação da MAG, foram retirados da base tempos menores ou iguais a zero. Nesta etapa, o total de observações removidas foi de 3.377.



4.1.3 Proporções de censura e resgate

Observou-se que 60,57% dos casos correspondem a resgates, enquanto 39,43% referem-se a censuras, que incluem tanto casos de censura propriamente ditos quanto saldamentos. Para uma descrição detalhada dos *status item contratado* classificados como censura, consulte o Apêndice B. Além disso, foram calculadas as proporções exatas para duas formas de desfecho dos produtos: resgate e saldamento. Essas proporções estão apresentadas na Figura 4.

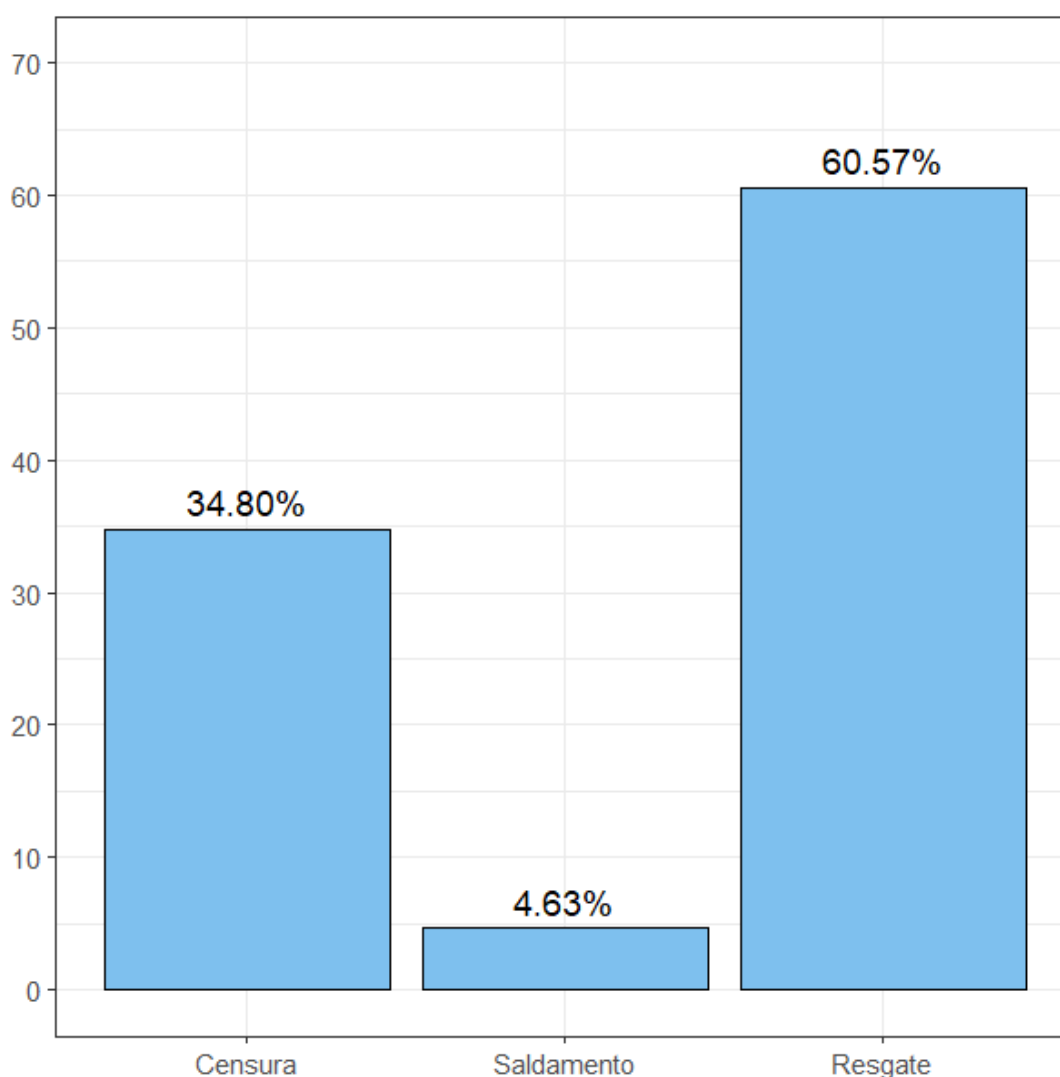


Figura 4: Desfecho Resgate: Proporções de censura e resgate (%).



4.1.4 Proporções por categoria do cliente

O gráfico apresentado na Figura 5 revela que 69,91% dos clientes estão na categoria *Cliente MAG*, enquanto 29,32% pertencem à categoria *MAG Blue*. A última coluna, que representa 0,78%, corresponde a clientes cuja categoria não foi especificada.

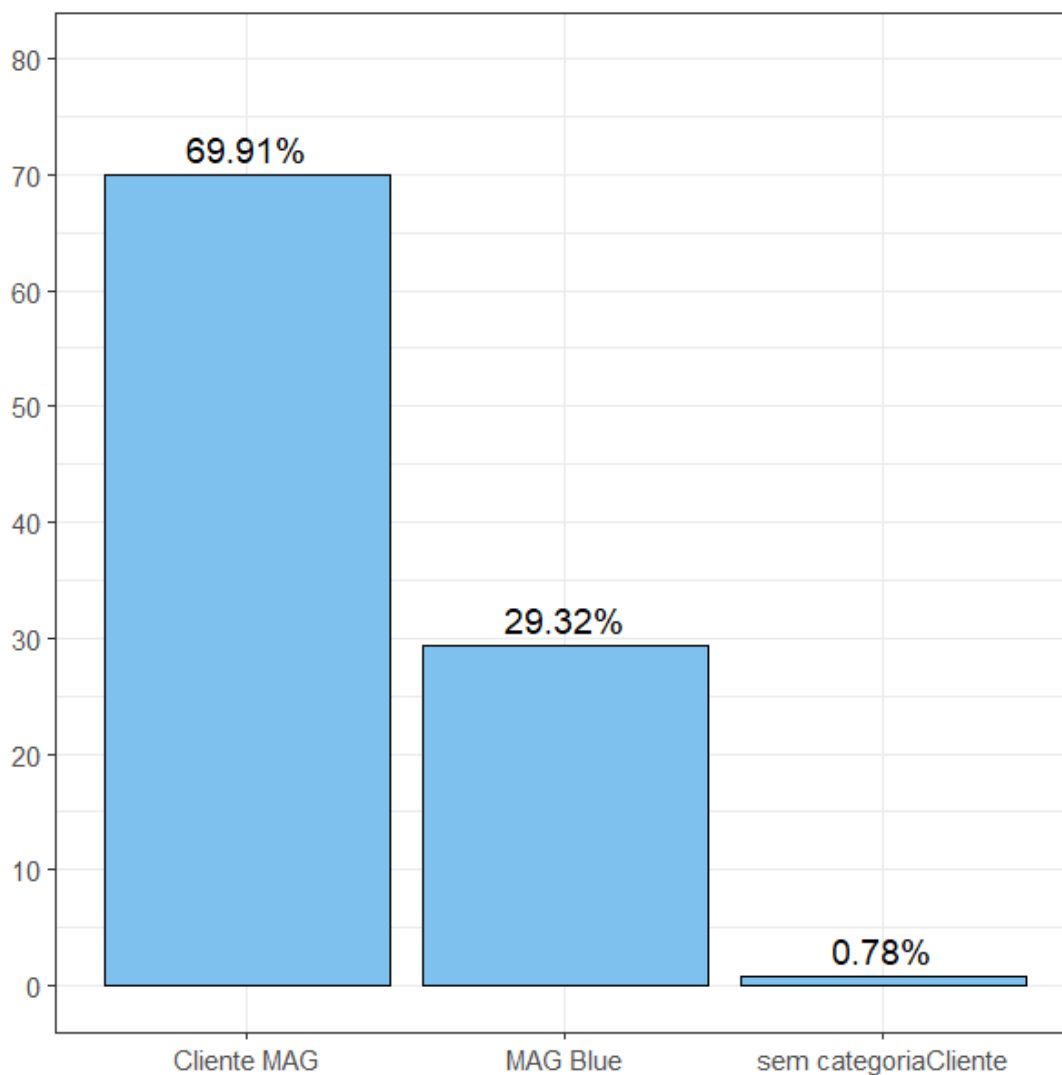


Figura 5: Desfecho Resgate: Proporções por categoria do cliente (%).



4.1.5 Proporções por forma de cobrança

O gráfico apresentado na Figura 6 mostra a distribuição percentual das diferentes formas de cobrança utilizadas. Observa-se que a forma de cobrança mais prevalente é o *DÉBITO CONTA*, representando 51,13% do total. Em segundo lugar, temos o *BOLETO*, com 24,94%, seguido pelo *CARTÃO DE CRÉDITO*, que corresponde a 13,79%. Outras formas, como *INSTITUÍDO*, *DESCONTO EM FOLHA*, *CONVÊNIO COBRANÇA* e *CONVÊNIO PARCERIA*, representam percentuais menores, com 6,47%, 3,04%, 0,63% e 0,00%, respectivamente.

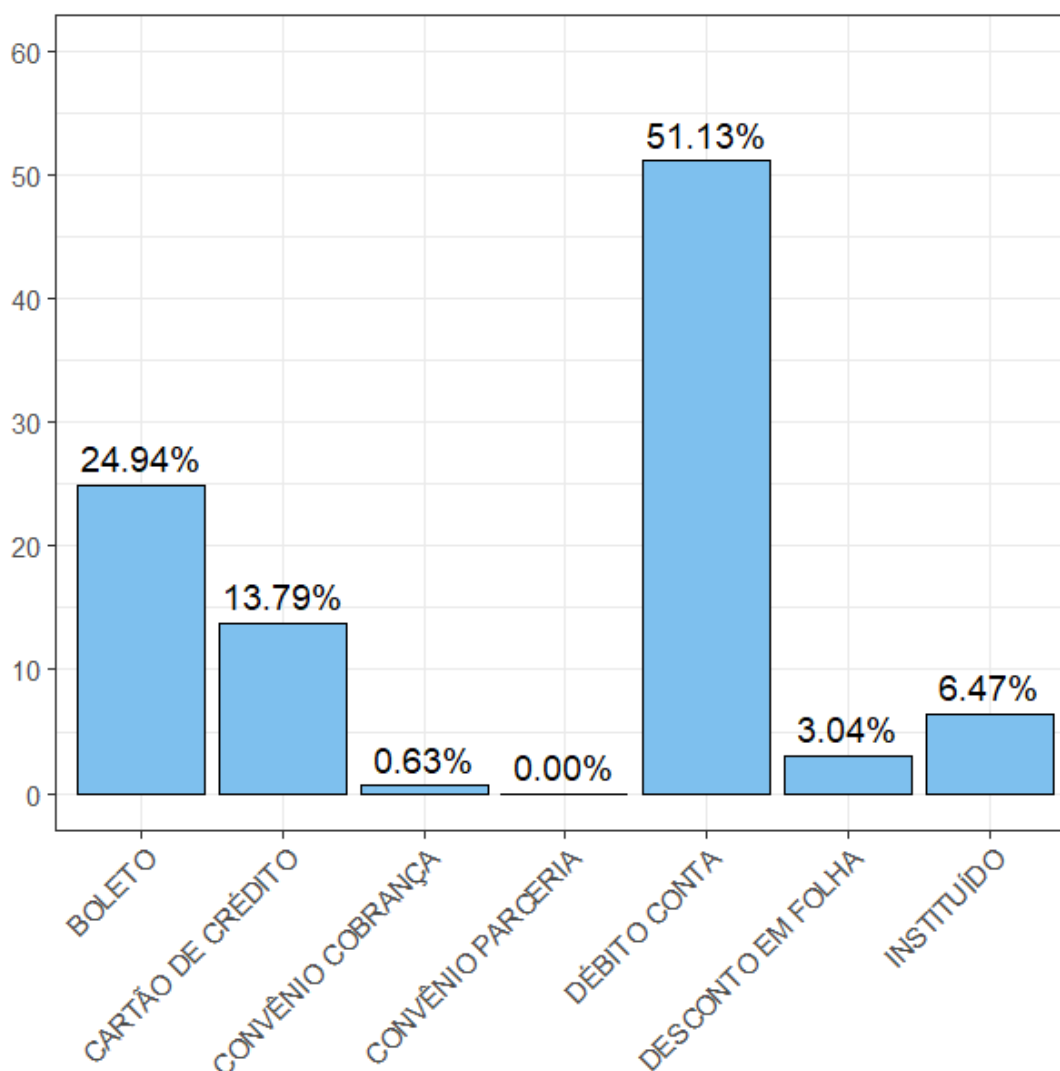


Figura 6: Desfecho Resgate: Proporções por forma de cobrança (%).



4.1.6 Proporções por tipo de cobertura

O gráfico apresentado na Figura 7 ilustra a distribuição percentual do tipo de cobertura. Observa-se que a cobertura por *MORTE* é a mais prevalente, representando 76,56% do total, enquanto a cobertura por *SOBREVIVÊNCIA* e *INVALIDEZ* representam 17,89% e 5,55%, respectivamente.

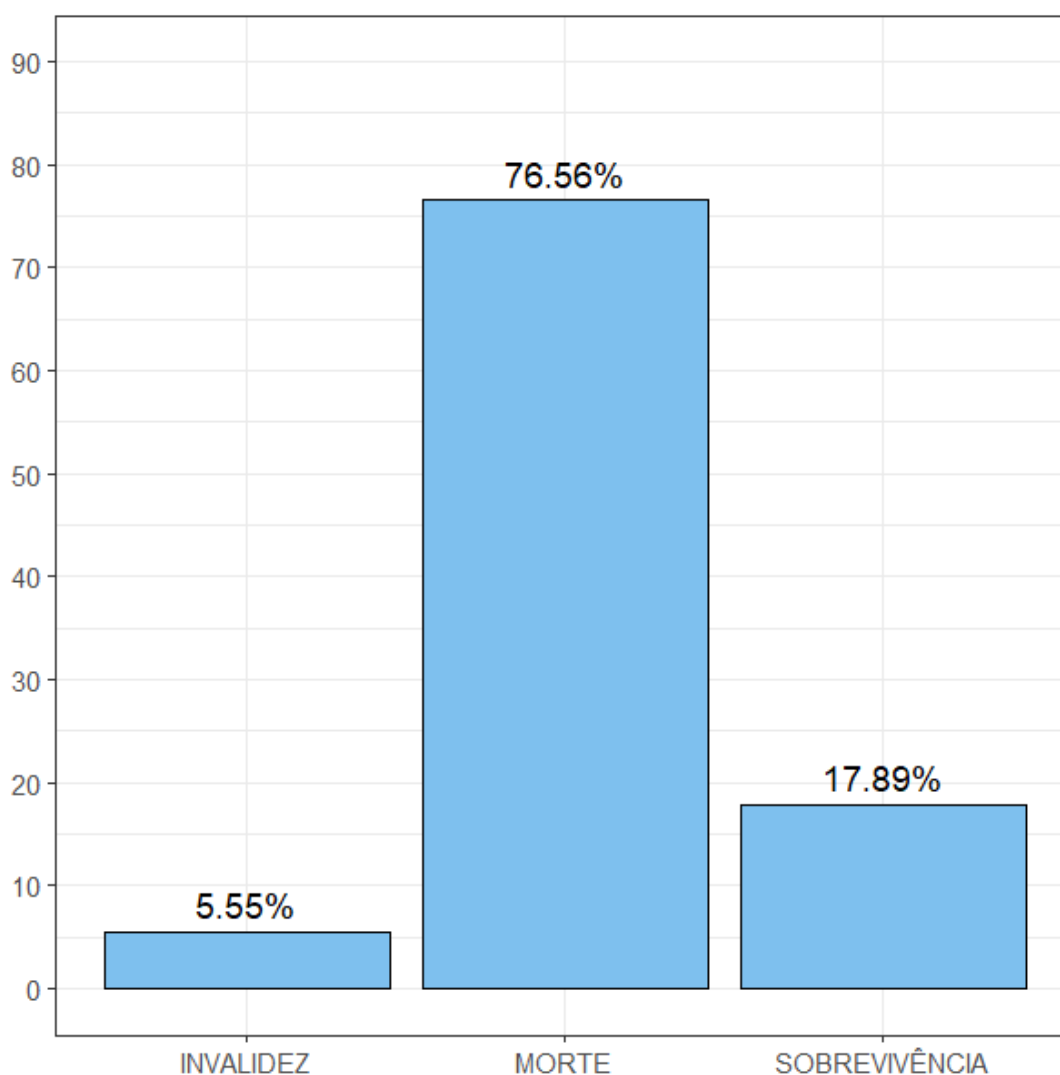


Figura 7: Desfecho Resgate: Proporções por tipo de cobertura (%).



4.1.7 Clusterização

A clusterização foi realizada com o objetivo de simplificar a análise, preservando as informações relevantes. As Figuras 1– 3 ilustram o comportamento de cada covariável antes da clusterização, evidenciando a variabilidade original e justificando a necessidade de redução. Uma descrição detalhada dos clusters formados para cada covariável está disponível no Apêndice A. Com o objetivo de reduzir o número de categorias e facilitar a modelagem, foram definidos os seguintes clusters:

- Categoria do Cliente: De 3 categorias, foram definidos 2 clusters.
- Forma de Cobrança: De 7 categorias, foram definidos 4 clusters.
- Tipo de Cobertura: De 3 categorias, foram definidos 2 clusters.

Durante a análise das combinações resultantes da clusterização das covariáveis, identificamos 14 cenários distintos, embora o total esperado fosse 16. A ausência de informações nos cenários específicos CC1-FC4-TC2 e CC2-FC4-TC2 resultou na inexistência desses cenários. A Figura 8 ilustra a função de risco observada para cada um desses cenários, destacando as diferenças e similaridades.

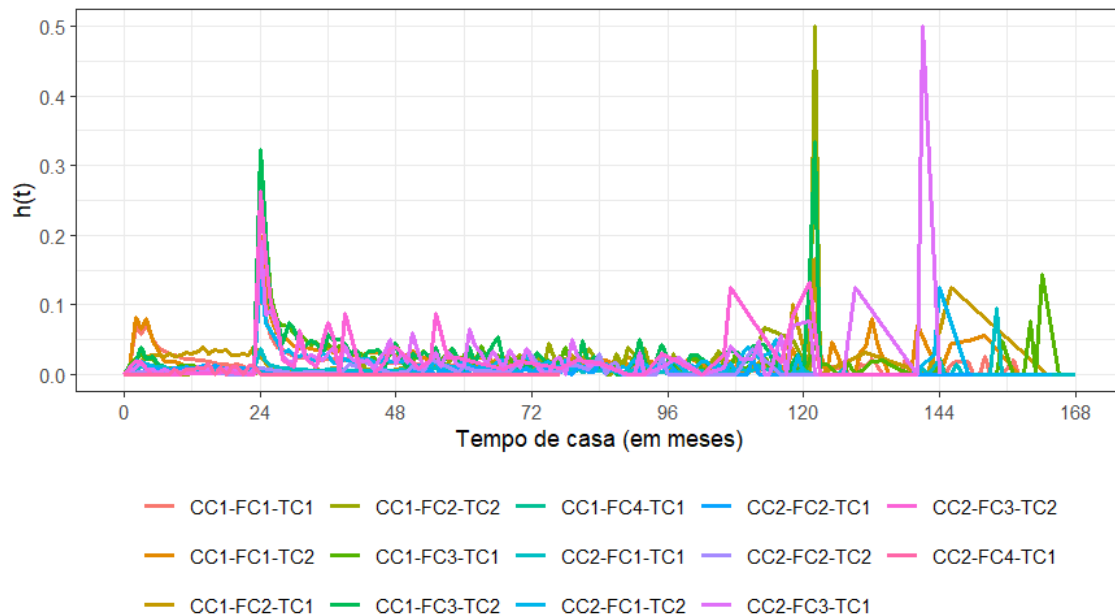


Figura 8: Desfecho Resgate: Função de risco observada.



Para reduzir a complexidade e facilitar a análise, esses cenários foram reorganizados com base na similaridade das funções de risco, resultando em 8 cenários finais. Os agrupamentos são detalhados a seguir:

- CC5-FC5-TC5 agrupa os cenários:
CC1-FC4-TC1 e CC2-FC4-TC1;
- CC6-FC6-TC6 agrupa os cenários:
CC1-FC2-TC2 e CC2-FC2-TC2;
- CC7-FC7-TC7 agrupa os cenários:
CC1-FC3-TC1; CC1-FC3-TC2; CC2-FC1-TC2; CC2-FC3-TC1 e CC2-FC3-TC2.

A Figura 9 apresenta a função de risco após o agrupamento, evidenciando a simplificação e a consolidação dos cenários.

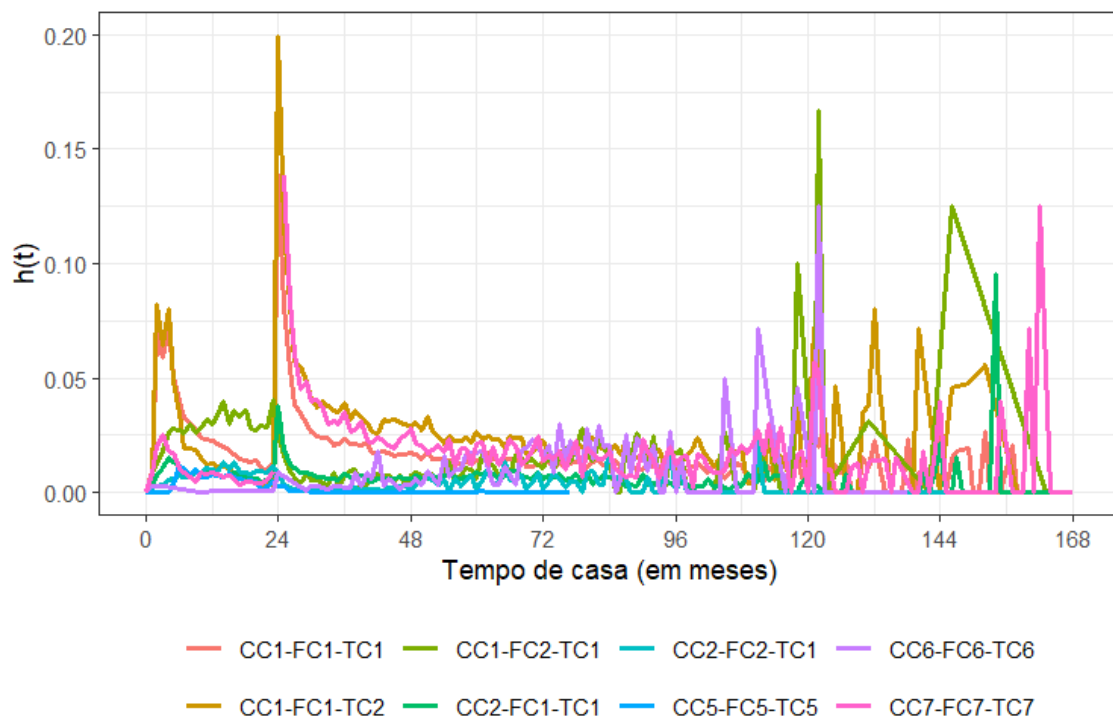


Figura 9: Desfecho Resgate: Função de risco observada - após o agrupamento dos cenários.

Esses três novos clusters oferecem uma base sólida para a análise, pois foram formados



a partir da identificação de combinações de clusters que apresentavam funções de risco semelhantes. Ao agrupar as combinações com comportamentos semelhantes, o número total de clusters foi reduzido, o que simplifica a análise e a interpretação dos resultados.

4.2 Aplicação do Modelo de Mistura

Nesta análise, foram consideradas as covariáveis *Categoria do Cliente*, *Forma de Cobrança* e *Tipo de Cobertura* na modelagem do tempo de persistência dos clientes. Os resultados são apresentados a seguir.

4.2.1 Cenários Avaliados

Para ajustar as curvas de persistência utilizando modelos de mistura, foram analisados diversos cenários de interesse, especificamente 8 cenários para o desfecho resgate. A configuração desses cenários é detalhada a seguir. Para uma descrição detalhada dos clusters associados a todas as covariáveis estudadas, consulte o apêndice [A](#).

- **Cenário 1:**

Cluster 1 (CC) + Cluster 1 (FC) + Cluster 1 (TC).

- **Cenário 2:**

Cluster 1 (CC) + Cluster 1 (FC) + Cluster 2 (TC).

- **Cenário 3:**

Cluster 1 (CC) + Cluster 2 (FC) + Cluster 1 (TC).

- **Cenário 4:**

Cluster 2 (CC) + Cluster 1 (FC) + Cluster 1 (TC).

- **Cenário 5:**

Cluster 2 (CC) + Cluster 2 (FC) + Cluster 1 (TC).

- **Cenário 6:**



Cluster 1 (CC) + Cluster 4 (FC) + Cluster 1 (TC) e

Cluster 2 (CC) + Cluster 4 (FC) + Cluster 1 (TC1).

- **Cenário 7:**

Cluster 1 (CC) + Cluster 2 (FC) + Cluster 2 (TC) e

Cluster 2 (CC) + Cluster 2 (FC) + Cluster 2 (TC).

- **Cenário 8:**

Outros.

4.2.2 Representatividade dos Cenários na Base

A Tabela 3 contém os números de observações de cada um dos cenários apresentados na subseção 4.2.1 na base de dados.

Tabela 3: Desfecho Resgate: Número de observações dos cenários na base

Cenário	Observações	Frequência (%)
CENÁRIO 1	91.577	41,89
CENÁRIO 2	33.036	15,11
CENÁRIO 3	9.922	4,54
CENÁRIO 4	38.804	17,75
CENÁRIO 5	19.867	9,09
CENÁRIO 6	14.152	6,47
CENÁRIO 7	356	0,16
CENÁRIO 8	10.885	4,98



4.2.3 Parâmetros Estimados

As Tabelas 4 e 5 apresentam as estimativas pontuais e intervalares obtidas para o modelo de sobrevivência com mistura Log-Normal.

Tabela 4: Desfecho Resgate: Modelo de sobrevivência com mistura Log-Normal

Parâmetro	Estimativa Pontual	IC (95%)
Peso da Mistura Componente 1 (η_1)	0,755	(0,751; 0,759)
Peso da Mistura Componente 2 (η_2)	0,150	(0,146; 0,154)
Peso da Mistura Componente 3 (η_3)	0,095	(0,093; 0,096)
Precisão Componente 1 (ϕ_1)	0,657	(0,646; 0,668)
Precisão Componente 2 (ϕ_2)	4,011	(3,888; 4,156)
Precisão Componente 3 (ϕ_3)	2.548,225	(2.463,002; 2.641,416)
Intercepto Componente 1 (β_1)	3,316	(3,301; 3,330)
Intercepto Componente 2 (β_2)	1,178	(1,166; 1,190)
Intercepto Componente 3 (β_3)	3,189	(3,189; 3,190)
Coeficiente CC2 Componente 1 ($\beta_{CC2,1}$)	1,515	(1,489; 1,539)
Coeficiente CC5 Componente 1 ($\beta_{CC5,1}$)	2,573	(-51,943; 51,001)
Coeficiente CC6 Componente 1 ($\beta_{CC6,1}$)	0,645	(-47,976; 54,284)
Coeficiente CC7 Componente 1 ($\beta_{CC7,1}$)	0,138	(-49,452; 50,151)
Coeficiente FC2 Componente 1 ($\beta_{FC2,1}$)	0,431	(0,395; 0,470)
Coeficiente FC5 Componente 1 ($\beta_{FC5,1}$)	2,019	(-48,393; 52,667)
Coeficiente FC6 Componente 1 ($\beta_{FC6,1}$)	-0,435	(-51,793; 51,014)
Coeficiente FC7 Componente 1 ($\beta_{FC7,1}$)	-0,842	(-49,461; 50,035)
Coeficiente TC2 Componente 1 ($\beta_{TC2,1}$)	-0,089	(-0,111; -0,067)
Coeficiente TC5 Componente 1 ($\beta_{TC5,1}$)	1,349	(-48,692; 52,514)
Coeficiente TC6 Componente 1 ($\beta_{TC6,1}$)	0,341	(-51,859; 51,863)
Coeficiente TC7 Componente 1 ($\beta_{TC7,1}$)	0,104	(-51,217; 52,309)



Tabela 5: Desfecho Resgate: Modelo de sobrevivência com mistura Log-Normal

Parâmetro	Estimativa Pontual	IC (95%)
Coeficiente CC2 Componente 2 ($\beta_{CC2,2}$)	0,608	(0,582; 0,635)
Coeficiente CC5 Componente 2 ($\beta_{CC5,2}$)	-0,273	(-48,826; 50,869)
Coeficiente CC6 Componente 2 ($\beta_{CC6,2}$)	1,346	(-51,531; 52,728)
Coeficiente CC7 Componente 2 ($\beta_{CC7,2}$)	-0,538	(-50,561; 49,483)
Coeficiente FC2 Componente 2 ($\beta_{FC2,2}$)	0,745	(0,709; 0,782)
Coeficiente FC5 Componente 2 ($\beta_{FC5,2}$)	0,109	(-51,344; 50,116)
Coeficiente FC6 Componente 2 ($\beta_{FC6,2}$)	1,080	(-50,036; 53,075)
Coeficiente FC7 Componente 2 ($\beta_{FC7,2}$)	-0,935	(-48,613; 50,619)
Coeficiente TC2 Componente 2 ($\beta_{TC2,2}$)	-0,084	(-0,103; -0,066)
Coeficiente TC5 Componente 2 ($\beta_{TC5,2}$)	1,555	(-46,753; 49,721)
Coeficiente TC6 Componente 2 ($\beta_{TC6,2}$)	1,830	(-49,081; 51,042)
Coeficiente TC7 Componente 2 ($\beta_{TC7,2}$)	-0,529	(-50,283; 51,535)
Coeficiente CC2 Componente 3 ($\beta_{CC2,3}$)	-0,001	(-0,002; 0)
Coeficiente CC5 Componente 3 ($\beta_{CC5,3}$)	-0,157	(-50,344; 51,797)
Coeficiente CC6 Componente 3 ($\beta_{CC6,3}$)	0,298	(-50,212; 52,006)
Coeficiente CC7 Componente 3 ($\beta_{CC7,3}$)	0,617	(-49,785; 49,039)
Coeficiente FC2 Componente 3 ($\beta_{FC2,3}$)	-0,547	(-0,549; -0,544)
Coeficiente FC5 Componente 3 ($\beta_{FC5,3}$)	1,263	(-48,101; 49,362)
Coeficiente FC6 Componente 3 ($\beta_{FC6,3}$)	0,954	(-47,099; 51,505)
Coeficiente FC7 Componente 3 ($\beta_{FC7,3}$)	-1,031	(-51,408; 50,348)
Coeficiente TC2 Componente 3 ($\beta_{TC2,3}$)	0	(-0,001; 0,001)
Coeficiente TC5 Componente 3 ($\beta_{TC5,3}$)	1,272	(-45,482; 49,049)
Coeficiente TC6 Componente 3 ($\beta_{TC6,3}$)	-0,181	(-49,288; 49,871)
Coeficiente TC7 Componente 3 ($\beta_{TC7,3}$)	1,008	(-50,036; 52,410)



4.2.4 Curvas de Persistência e Taxas de Falha

Nesta seção, apresentamos gráficos comparando as curvas de persistência empíricas estimadas via Kaplan-Meier (KM) (linha sólida em preto) com as ajustadas pelo modelo de sobrevivência com mistura Log-Normal (MSMLN) (linha sólida em azul). Ambas as curvas foram construídas com base nos dados completos a partir de 2010, permitindo avaliar a adequação do modelo ao comportamento de persistência observado. A seleção da curva final foi realizada com base em uma análise de sensibilidade, na qual foi escolhido um modelo com três componentes de mistura, que apresentou o melhor ajuste, priorizando a qualidade visual em relação às métricas quantitativas.

Além das curvas de persistência, também analisamos as taxas de falha, que fornecem uma medida do risco instantâneo de resgate por parte do cliente. Em um tempo fixo de interesse, a taxa de falha indica o risco do cliente resgatar a inscrição naquele exato momento.

Assim, apresentaremos, nas Figuras 10 a 17, as curvas de persistência e as curvas de taxas de falha empíricas e ajustadas pelo MSMLN para cada um dos cenários estudados.

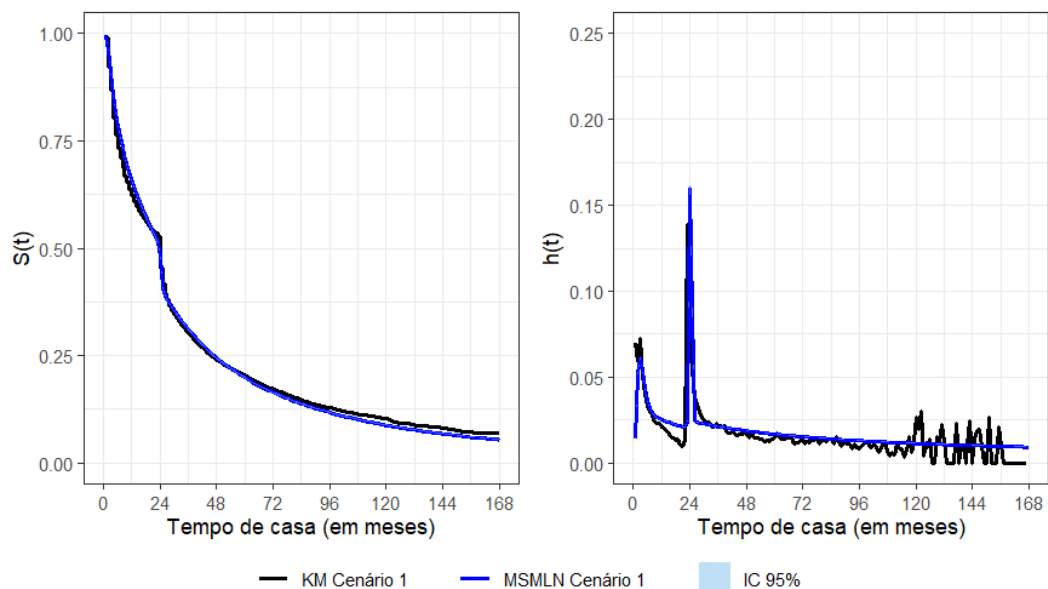


Figura 10: Desfecho Resgate: Cenário 1 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

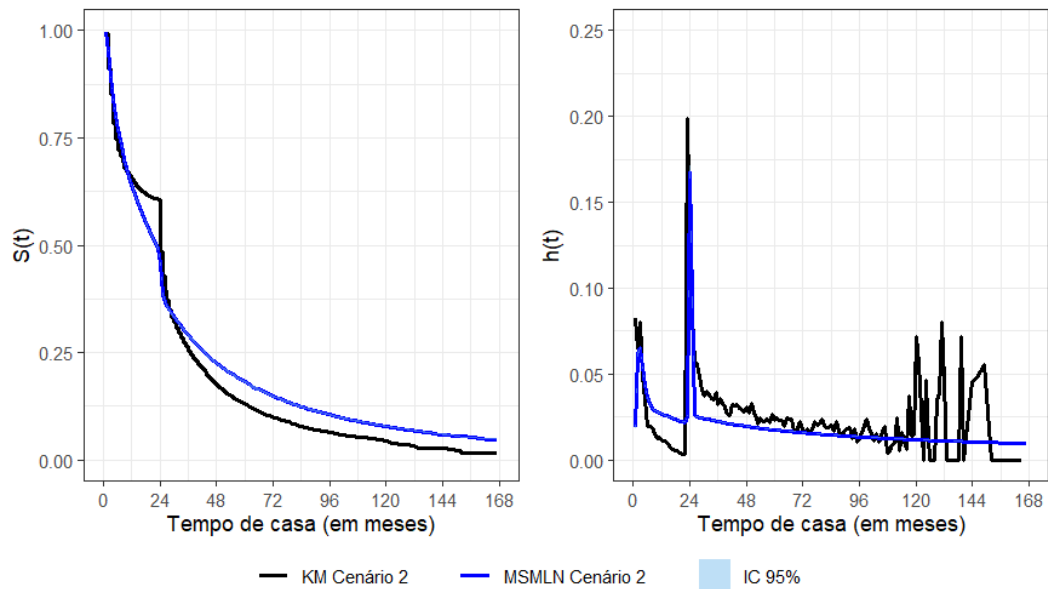


Figura 11: Desfecho Resgate: Cenário 2 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

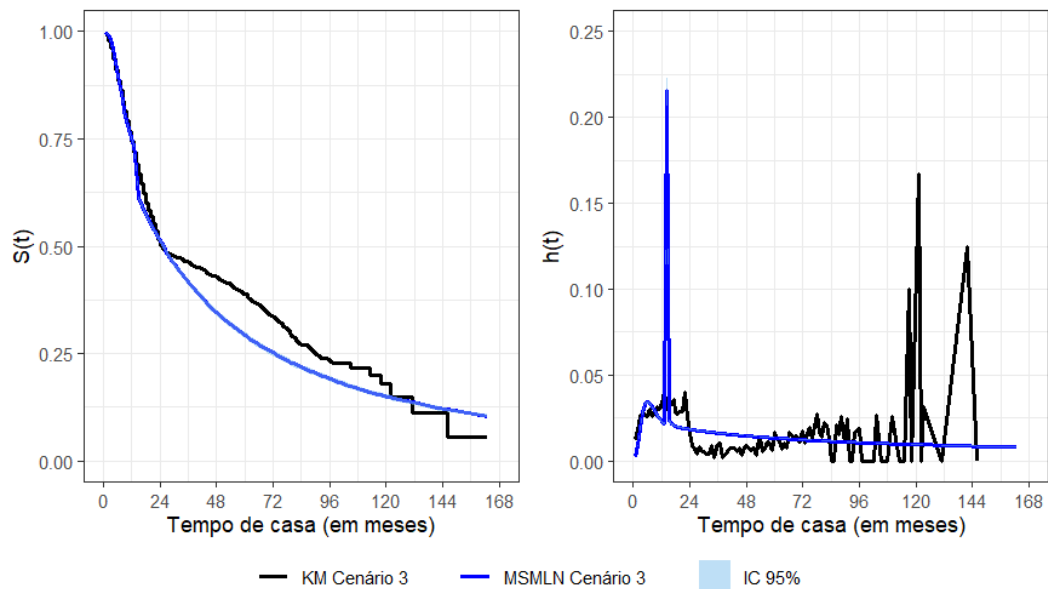


Figura 12: Desfecho Resgate: Cenário 3 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

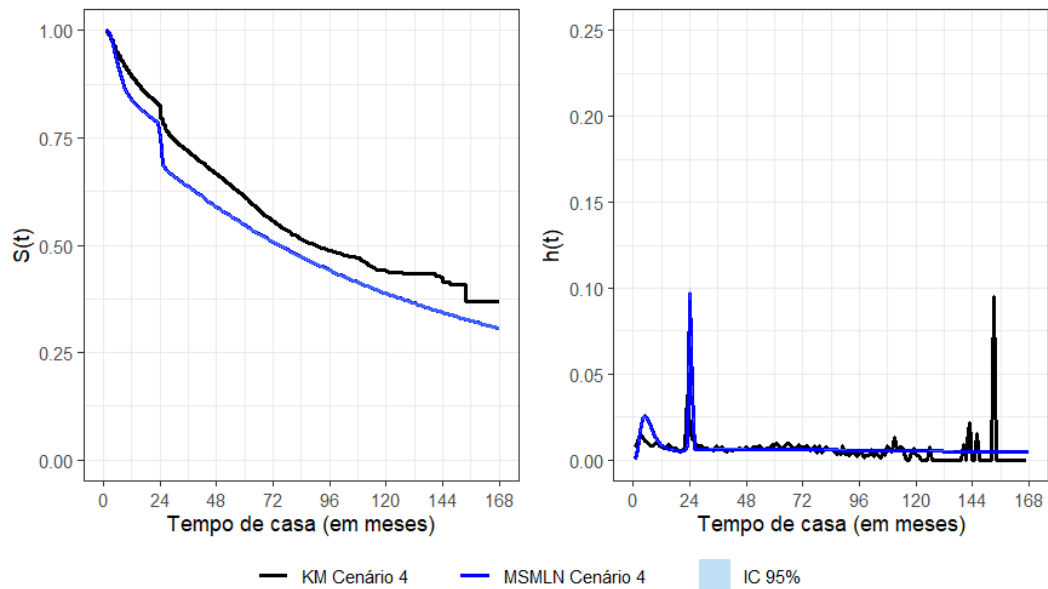


Figura 13: Desfecho Resgate: Cenário 4 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

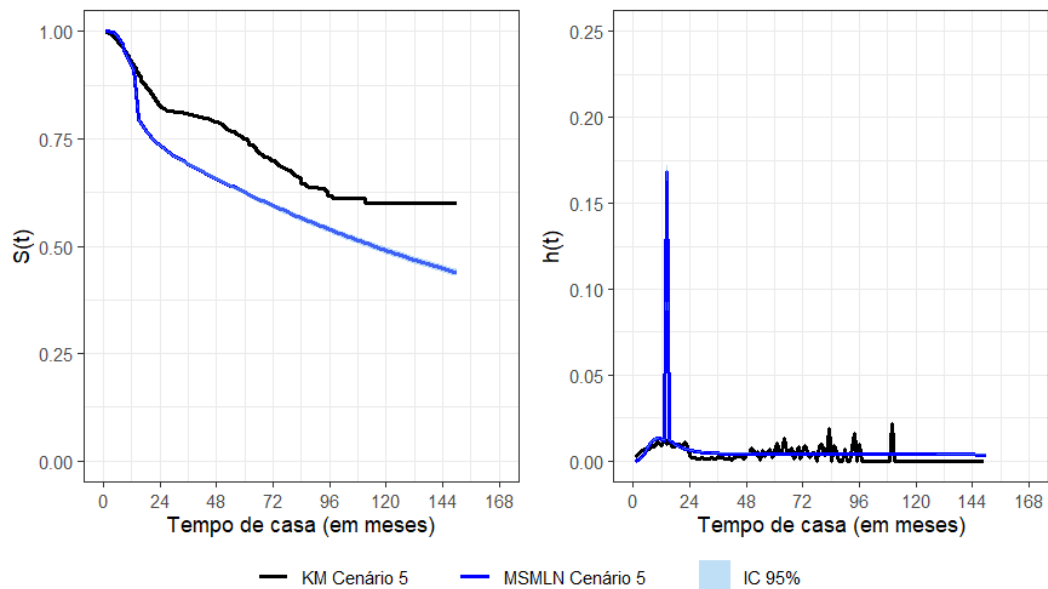


Figura 14: Desfecho Resgate: Cenário 5 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

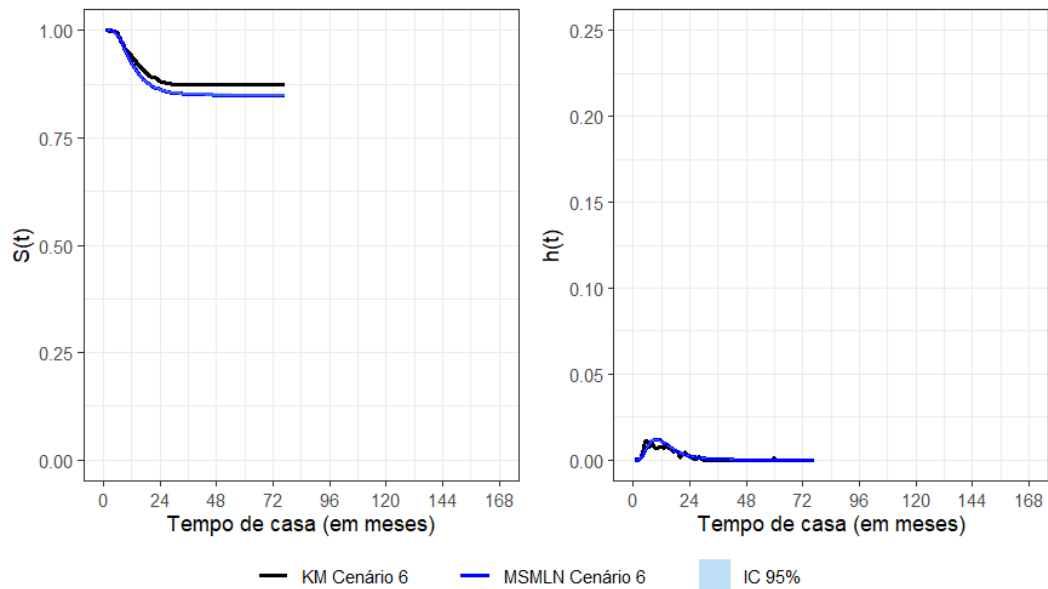


Figura 15: Desfecho Resgate: Cenário 6 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

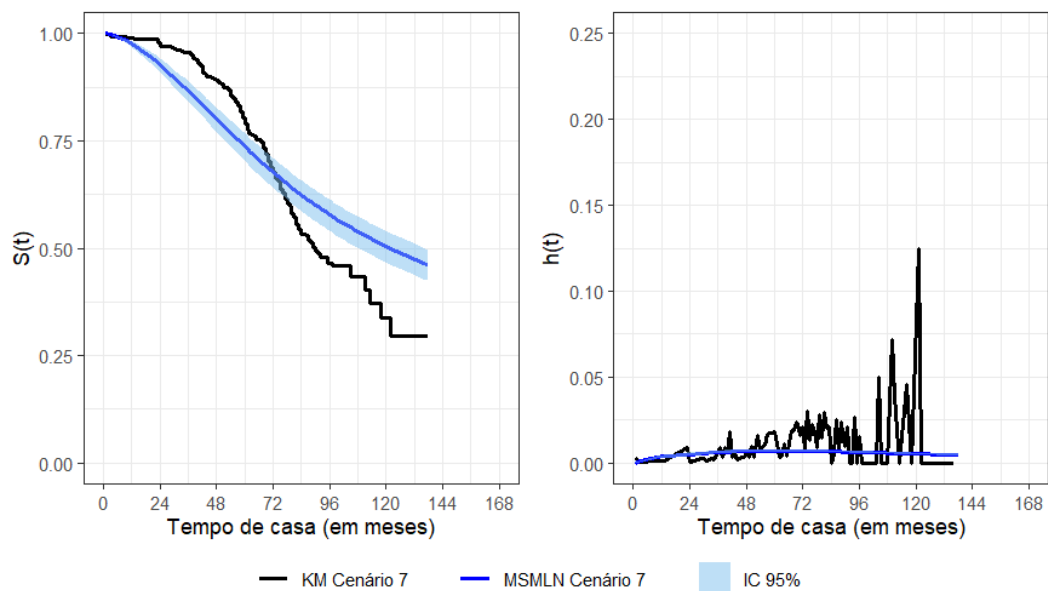


Figura 16: Desfecho Resgate: Cenário 7 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

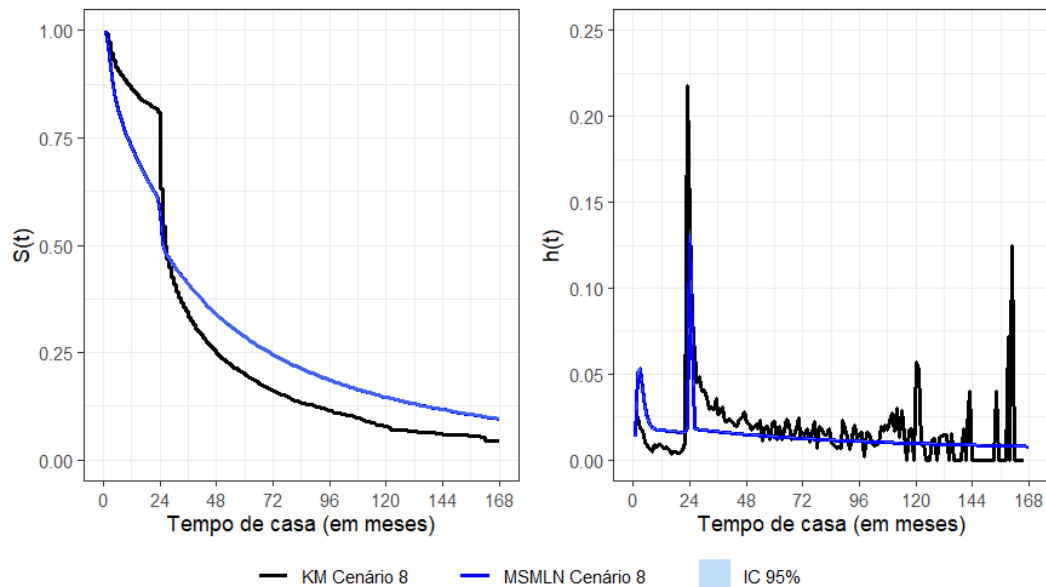


Figura 17: Desfecho Resgate: Cenário 8 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica obtida por Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

Observa-se, pelas Figuras 10 a 17, que o modelo de sobrevivência com mistura Log-Normal (MSMLN) apresenta um bom ajuste aos dados na maioria dos cenários, capturando as características observadas nas curvas empíricas. No entanto, algumas exceções são identificadas nos seguintes casos: o Cenário 7, ilustrado na Figura 16, e os Cenários 3 e 5, ilustrados nas Figuras 12 e 14, respectivamente. No Cenário 7, o ajuste é menos preciso devido ao menor número de observações. Nos Cenários 3 e 5, os picos nas taxas de falha podem ser resultado da modelagem conjunta com outros cenários, embora não estejam presentes nos dados empíricos desses cenários, prejudicando assim o ajuste.



4.2.5 Comparações das Curvas de Persistência e Taxas de Falha para os cenários

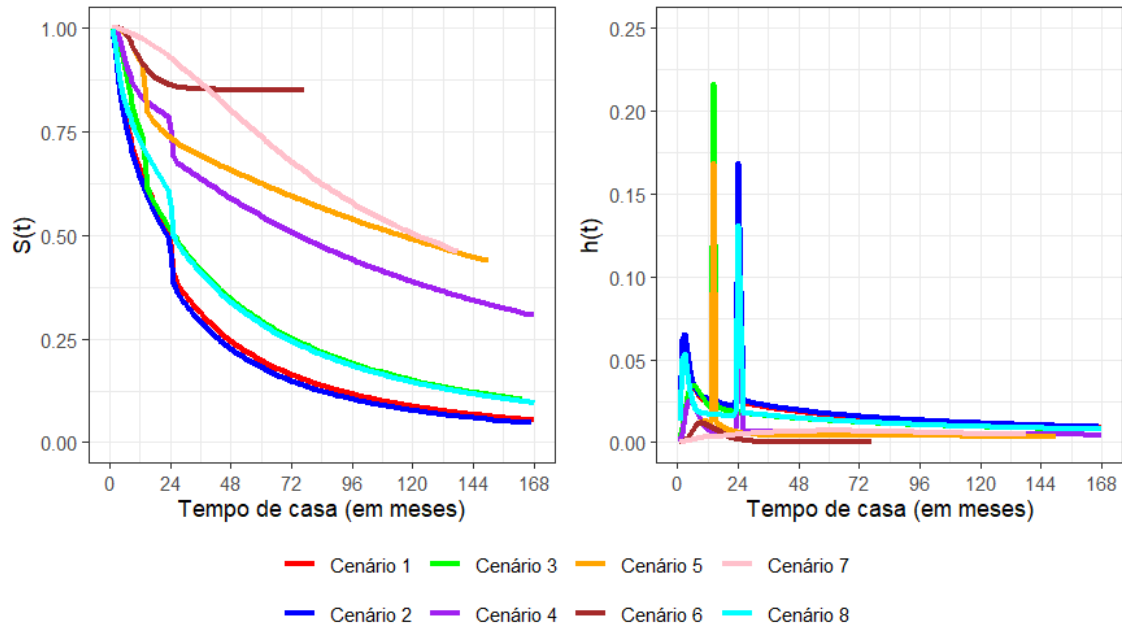


Figura 18: Desfecho Resgate: Curvas de persistência e taxas de falha ajustadas pelo MSMLN para todos os cenários.

Esta seção compara, a partir das curvas apresentadas na Figura 18, as persistências e taxas de falha observadas para diferentes cenários. Como resultado, podemos avaliar os riscos de resgate instantâneos para clientes de diferentes cenários de interesse.



Cenário 1 versus Cenário 3

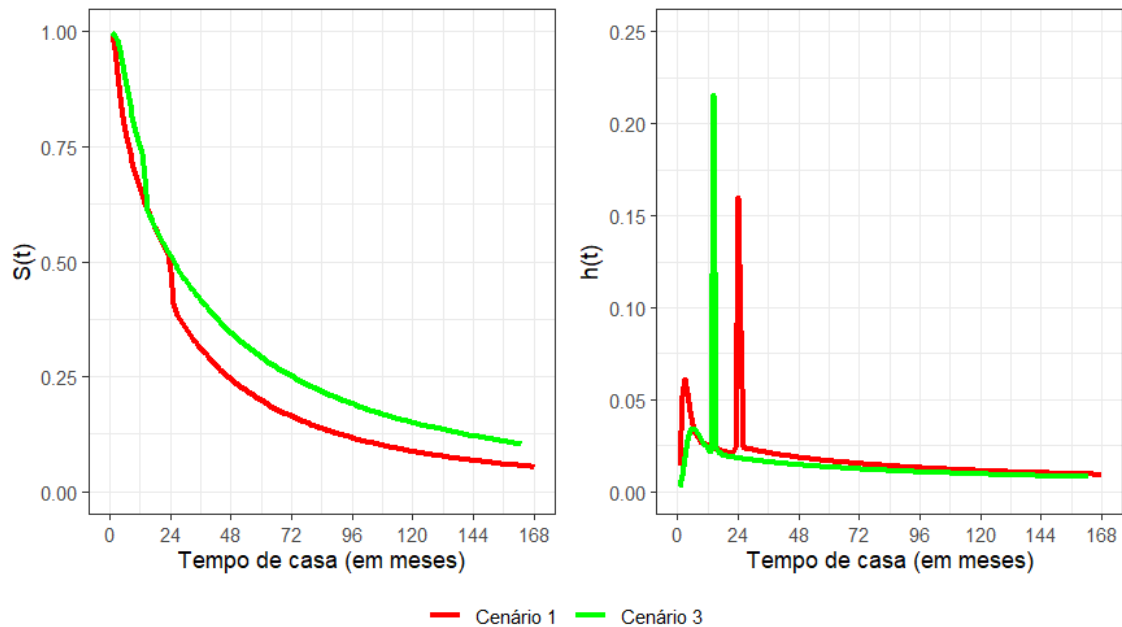


Figura 19: Desfecho Resgate: Curvas de persistência e taxas de falha ajustadas pelo MSMLN para os Cenários 1 e 3.

Conforme descrição apresentada na subseção 4.2.1, é possível notar que a diferença observada entre os Cenários 1 e 3 se dá apenas pela variável *Forma de Cobrança*. Enquanto no Cenário 1 os clientes são do Cluster 1 dessa variável (*BOLETO*, *CONVÊNIO COBRANÇA* e *DÉBITO CONTA*), no Cenário 3 os clientes são do Cluster 2 (*CARTÃO DE CRÉDITO* e *CONVÊNIO PARCERIA*). A partir disso, é interessante comparar as curvas de persistência e taxas de falha de ambos os cenários para verificar qual o impacto que essa variável em especial pode gerar no tempo de permanência do cliente.

Como mostra a Figura 19, os clientes do Cenário 3 tendem a persistir mais do que clientes do Cenário 1. Dessa forma, podemos concluir que, para clientes do Cluster 1 de *Categoria do Cliente* e do Cluster 1 de *Tipo de Cobertura*, aqueles que pertencem ao Cluster 2 de *Forma de Cobrança* tendem a permanecer na empresa por mais tempo do que aqueles que pertencem ao Cluster 1. Ao olhar a taxa de risco, temos que, ao longo de quase todo o período observado, o risco de resgate dos clientes do Cenário 3 é menor do que o dos clientes do Cenário 1, com a exceção de um pico no tempo 14 para clientes



do Cenário 3. Esse pico pode ter ocorrido porque o MSMLN foi influenciado pelos picos presentes em outros cenários, resultando em um pico que não existe.

Se fizermos a comparação Cenário 4 *versus* Cenário 5, na qual a única diferença também se dá apenas pela variável *Forma de Cobrança*, chegaremos às mesmas conclusões, de que os clientes do Cluster 2 dessa variável persistem mais do que clientes do Cluster 1.

Sendo assim, visando a captação de novos clientes que persistam por mais tempo na empresa, pode ser interessante priorizar a forma de cobrança do Cluster 2 em relação aos do Cluster 1, visto que, de uma forma geral, esse Cluster vai apresentar melhores resultados.

Cenário 1 *versus* Cenário 2

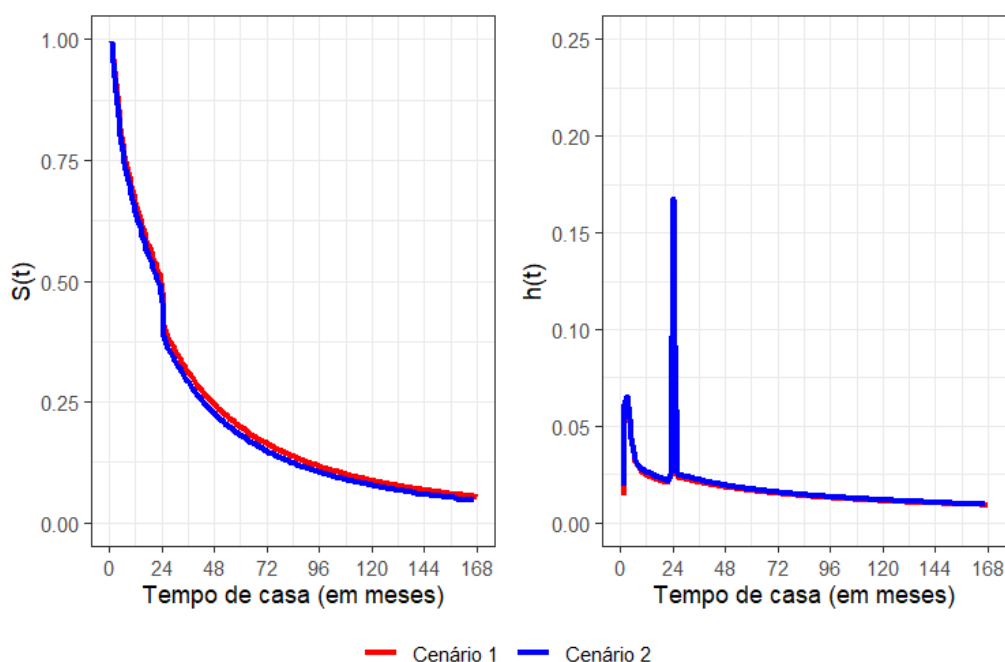


Figura 20: Desfecho Resgate: Curvas de persistência e taxas de falha ajustadas pelo MSMLN para os Cenários 1 e 2.

Conforme descrição apresentada na subseção 4.2.1, é possível notar que a diferença observada entre os Cenários 1 e 2 se dá apenas pela variável *Tipo de Cobertura*. Enquanto no Cenário 1 os clientes são do Cluster 1 dessa variável (*INVALIDEZ* e *MORTE*), no



Cenário 2 os clientes são do Cluster 2 (*SOBREVIVÊNCIA*). A partir disso, é interessante comparar as curvas de persistência e taxas de falha de ambos os cenários para verificar qual o impacto que essa variável em especial pode gerar no tempo de permanência do cliente.

Como mostra a Figura 20, as curvas de ambos os cenários são praticamente idênticas, com pequenas diferenças visíveis em alguns momentos. Isso sugere que a persistência dos clientes é muito semelhante nos dois cenários, indicando que a variável *Tipo de Cobertura* que diferencia esses cenários não tem um impacto significativo na probabilidade de permanência ao longo do tempo.

4.3 Probabilidades de Persistência

Na presente subseção, serão calculadas algumas probabilidades de persistência de interesse para o desfecho resgate. A partir de características do cliente, diremos com que probabilidade este cliente irá persistir por um certo tempo na empresa.

4.3.1 Exemplo 1

Um cliente novo na empresa é classificado como um Cliente MAG, com forma de cobrança Boleto e tipo de cobertura Morte. Qual a probabilidade desse cliente permanecer na empresa por 1 ano?

De acordo com o apêndice A, ao final do relatório, onde encontra-se uma descrição dos Clusters para todas as variáveis estudadas, podemos concluir que o cliente em questão pertence ao Cluster 1 de *Categoria do Cliente*, ao Cluster 1 de *Forma de Cobrança* e ao Cluster 1 de *Tipo de Cobertura*. Pela descrição dos cenários em 4.2.1, podemos verificar que o cliente em questão está inserido no que chamamos de Cenário 1.

Utilizando-se do modelo ajustado, encontramos uma estimativa $\widehat{S}(t)$ para a função de sobrevivência no tempo de interesse, no caso $t = 12$ meses. Dessa forma, encontramos que $\widehat{S}(12) = 0,6612$, e portanto a probabilidade do cliente em questão permanecer na empresa por 1 ano é de aproximadamente 66,12%.



4.3.2 Exemplo 2

Suponha agora que passaram-se 6 meses e que o mesmo cliente do Exemplo 1 ainda permanece na empresa. Queremos atualizar a probabilidade desse cliente permanecer por 1 ano na empresa, dado que ele já permaneceu por 6 meses. Para isso, faremos:

$$P(T > 12 \mid T > 6) = \frac{P((T > 12) \cap (T > 6))}{P(T > 6)} = \frac{P(T > 12)}{P(T > 6)} = \frac{S(12)}{S(6)}. \quad (16)$$

Já calculamos no Exemplo 1 o valor de $\widehat{S(12)}$. Precisamos agora calcular $\widehat{S(6)}$, de forma análoga ao que foi feito para calcular $\widehat{S(12)}$, porém com $t = 6$ meses. Ao fazer os cálculos, obtemos que $\widehat{S(6)} = 0,7842$. Dessa forma, temos que:

$$P(T > 12 \mid T > 6) = \frac{S(12)}{S(6)} = \frac{0,6612}{0,7842} = 0,8432. \quad (17)$$

Isto é, após permanecer por 6 meses, a probabilidade do cliente em questão persistir na empresa por 1 ano aumentou de 66,12% para 84,32%. Esse aumento é coerente, uma vez que, por já haver permanecido por algum tempo como cliente da empresa, é mais provável que ele seja um cliente fiel. Além disso, agora ele precisa permanecer por menos tempo na empresa para completar 1 ano na mesma, isto é, ele só precisa permanecer mais 6 meses, e não mais 12 meses, uma vez que 6 meses já se passaram.

4.3.3 Exemplo 3

Um novo cliente da empresa possui exatamente as mesmas características do cliente estudado nos dois exemplos anteriores, porém este novo cliente é classificado como Cartão de Crédito, ao contrário do cliente anterior, que era classificado como Boleto. Desejamos saber com que probabilidade esse novo cliente permanecerá na empresa pelo período de 1 ano.

Pelas informações fornecidas, temos que este novo cliente pertencente ao Cluster 1 de *Categoria do Cliente*, ao Cluster 2 de *Forma de Cobrança* e ao Cluster 1 de *Tipo de Cobertura*. Pela descrição dos cenários em 4.2.1, podemos verificar que o cliente em questão está inserido no que chamamos de Cenário 3.



Mais uma vez utilizando-se do modelo ajustado, obtemos que $\widehat{S(12)} = 0,7535$. Ou seja, a probabilidade do cliente em questão permanecer na empresa por 1 ano é de aproximadamente 75,35%, e portanto maior do que a probabilidade do cliente do Exemplo 1 permanecer pelo mesmo tempo (66,12%). Tal resultado é coerente, uma vez que, conforme verificado na subseção 4.2.5 e ilustrado na Figura 19, os clientes que utilizam Cartão de Crédito, em geral, apresentam maior persistência na empresa do que aqueles que optam pelo Boleto.



5 Análise do desfecho Saldamento

5.1 Base de dados

Nesta seção, será apresentada a descrição do conjunto de dados fornecido pela empresa, destacando os agrupamentos realizadas nas categorias das covariáveis, bem como nas remoções efetuadas na base de dados. É importante ressaltar que a base saldamento difere da base resgate no que diz respeito ao *Tipo de Cobertura*. Enquanto na base resgate há a categoria Sobrevivência na covariável *Tipo de Cobertura*, essa categoria não está presente na base saldamento.

5.1.1 Descrição dos dados

A Tabela 6 apresenta um resumo das covariáveis utilizadas para descrever o processo de persistência, onde a variável de interesse é o tempo (em meses) até o desfecho saldamento. Entre as três covariáveis, duas são do tipo categórica e uma é do tipo binária. Para simplificar a análise e facilitar a interpretação dos resultados, a única covariável que passou pelo processo de clusterização foi a *Forma de Cobrança*. Esta escolha foi motivada pelo elevado número de categorias presentes nessa covariável, o que poderia tornar complexa a interpretação dos resultados da modelagem.

Tabela 6: Desfecho Saldamento: Descrição das covariáveis.

Covariável	Tipo	Suporte
Categoria do Cliente	Categórica	$\{1, 2, 3\}$
Forma de Cobrança	Categórica	$\{1, \dots, 7\}$
Tipo de Cobertura	Binária	$\{1, 2\}$

As Figuras 21– 23 ilustram as funções de sobrevivência e risco associadas às covariáveis *Categoria do Cliente*, *Forma de Cobrança* e *Tipo de Cobertura*, respectivamente. Essas representações permitem uma análise visual das diferenças entre as categorias, sendo cruciais para compreender as características das covariáveis estudadas e justificar a necessidade de clusterização.

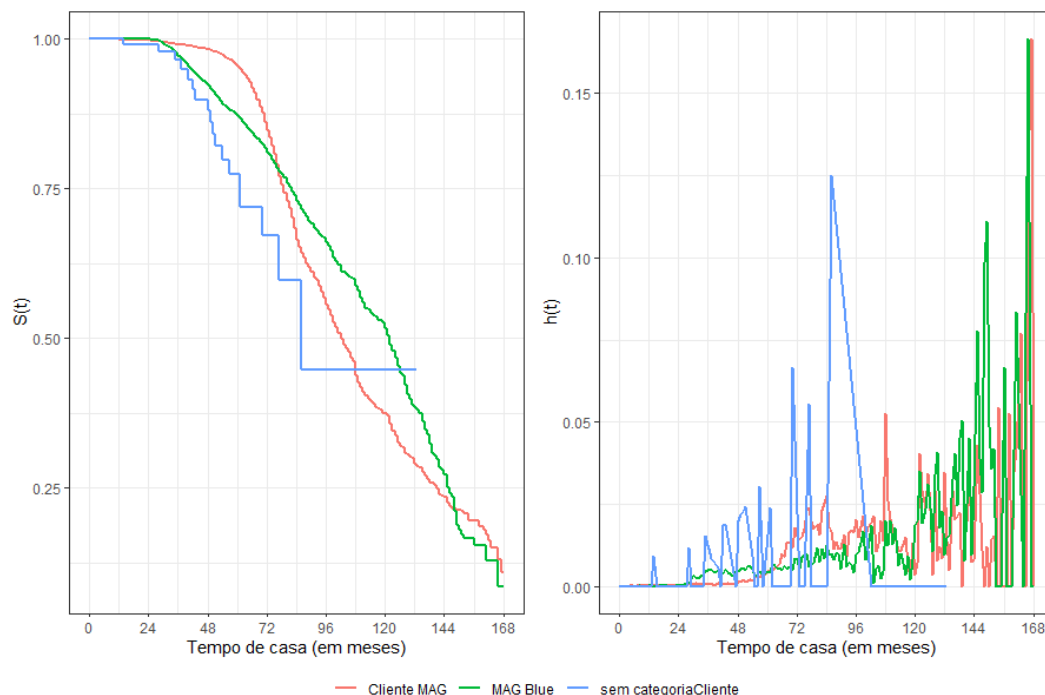


Figura 21: Desfecho Saldamento: Função de Sobrevivência e Risco - Categoria do Cliente.

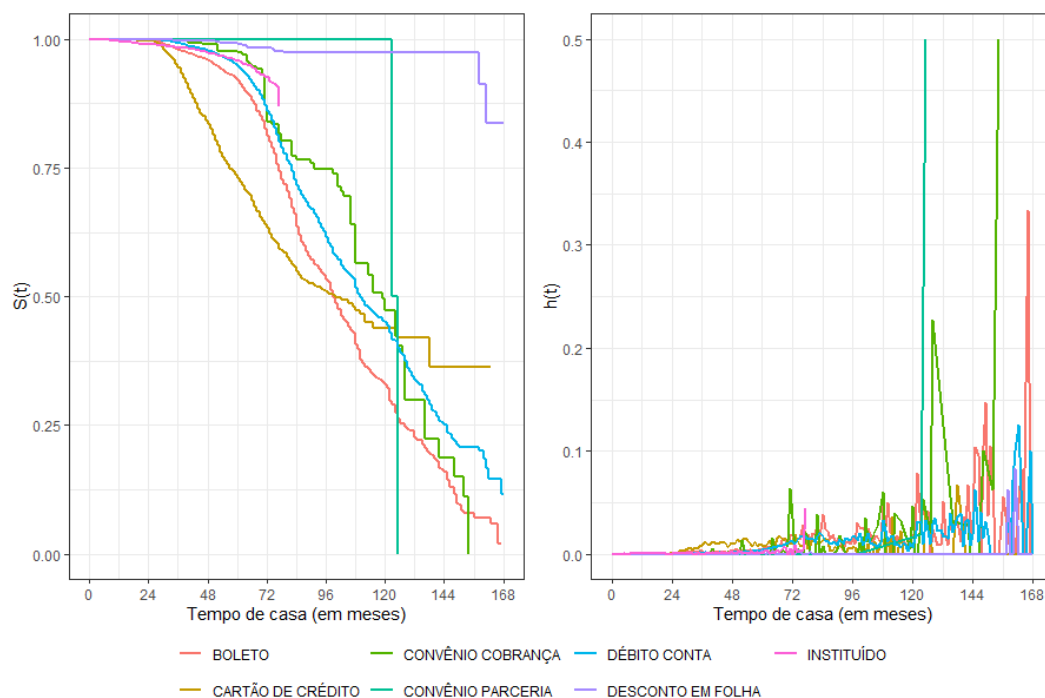


Figura 22: Desfecho Saldamento: Função de Sobrevivência e Risco - Forma de Cobrança.

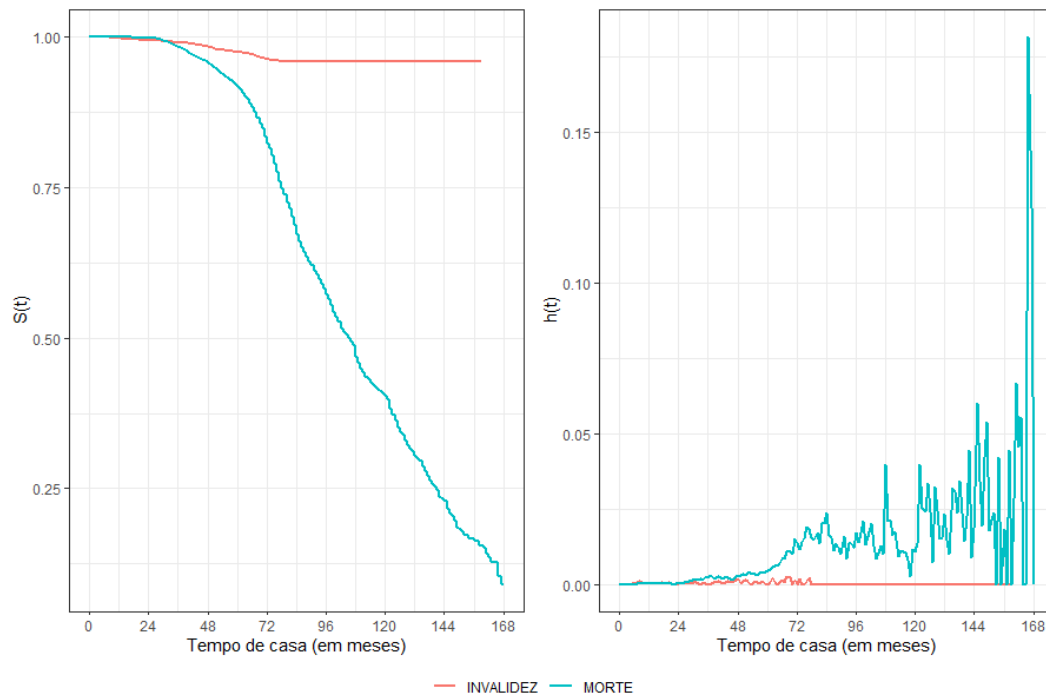


Figura 23: Desfecho Saldamento: Função de Sobrevivência e Risco - Tipo de Cobertura.

Durante o estudo, foram analisadas outras covariáveis como *Sexo*, *Periodicidade de Cobrança* e *Cobertura Agrupamento*. No entanto, essas covariáveis não foram consideradas no estudo, pois a análise dos dados indicou que não apresentavam significância estatística relevante para o modelo.

5.1.2 Remoções

Na base saldamento, as remoções são as mesmas descritas na Tabela 2 da subseção 4.1.2. Essas remoções foram realizadas na base original, antes da separação entre os desfechos resgate e saldamento.

Outras remoções: Tempos até o desfecho

Na base saldamento, os tempos removidos são os mesmos descritos na base resgate, na subseção 4.1.2.



5.1.3 Proporções de censura e saldamento

Observou-se que 5,64% dos casos correspondem a saldamentos, enquanto 94,36% referem-se a censuras, que incluem tanto casos de censura propriamente ditos quanto resgates. Para uma descrição detalhada dos *status item contratado* classificados como censura, consulte o Apêndice B. Além disso, foram calculadas as proporções exatas para duas formas de desfechos dos produtos: resgate e saldamento. Essas proporções estão apresentadas na Figura 24.

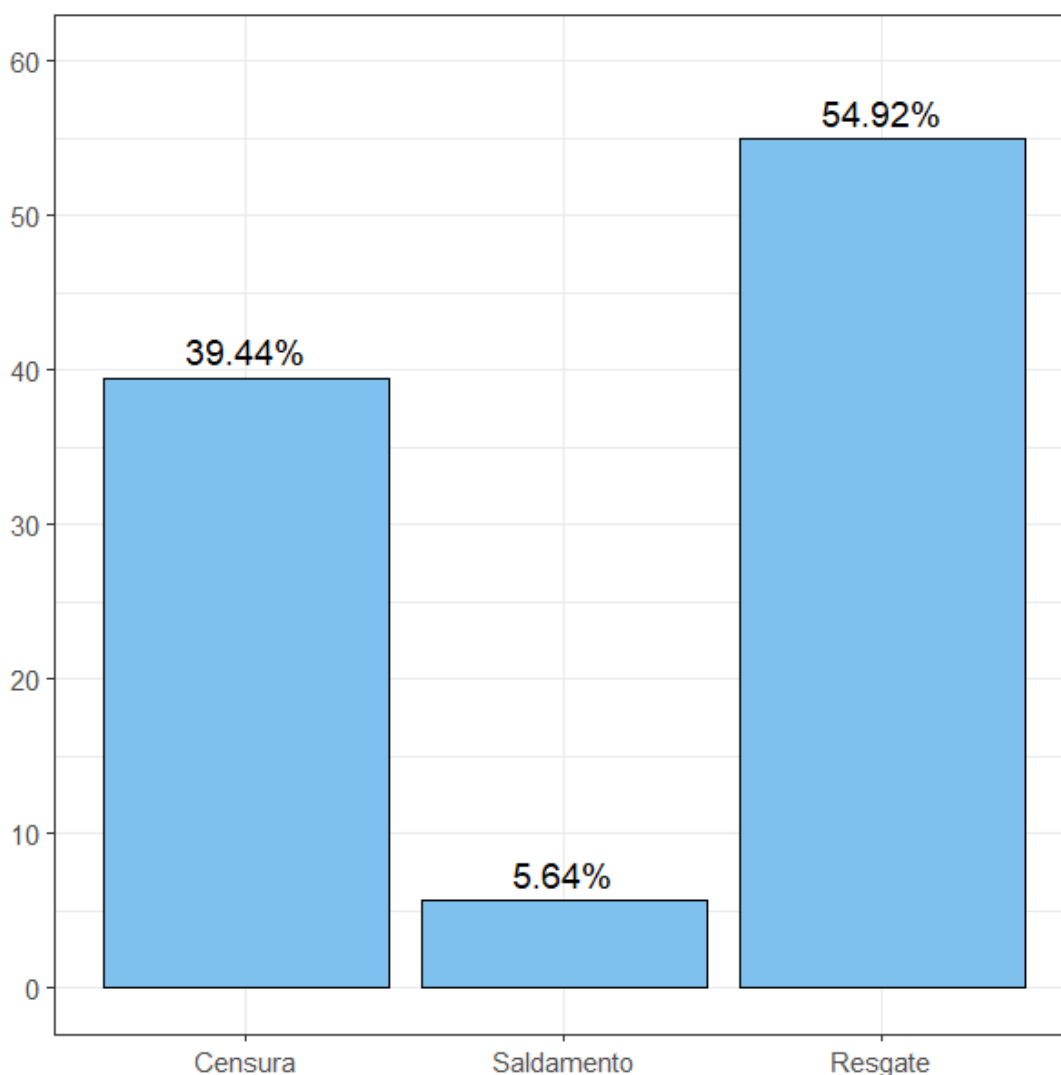


Figura 24: Desfecho Saldamento: Proporções de censura e saldamento (%).



5.1.4 Proporções por categoria do cliente

A Figura 25 revela que 65,85% dos clientes estão na categoria *Cliente MAG*, enquanto 33,22% pertencem à categoria *MAG Blue*. A última coluna, que representa 0,93%, corresponde a clientes cuja categoria não foi especificada.

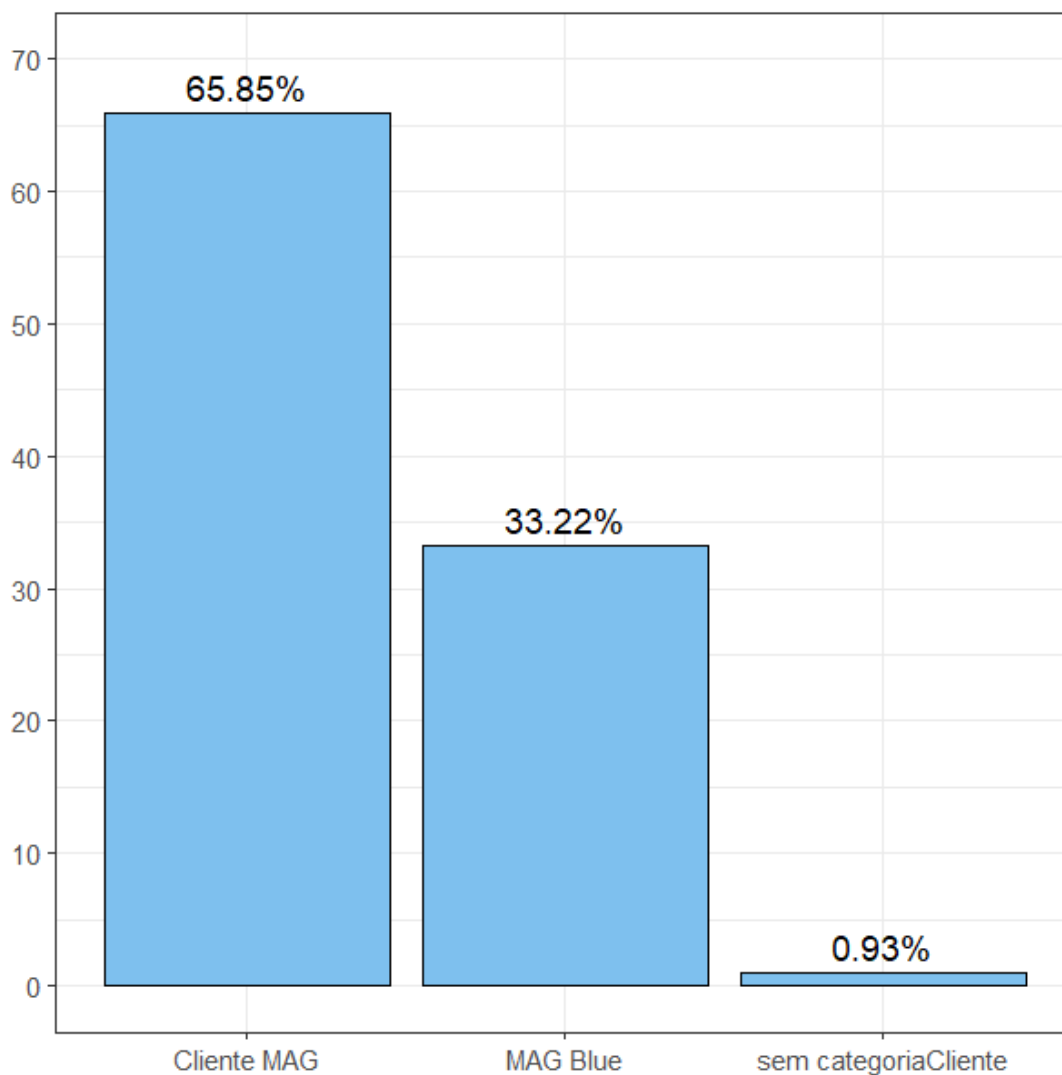


Figura 25: Desfecho Saldamento: Proporções por categoria (%)



5.1.5 Proporções por forma de cobrança

O gráfico apresentado na Figura 26 mostra a distribuição percentual das diferentes formas de cobrança utilizadas. Observa-se que a forma de cobrança mais prevalente é o *DÉBITO CONTA*, representando 48,40% do total. Em segundo lugar, temos o *BOLETO*, com 23,64%, seguido pelo *CARTÃO DE CRÉDITO*, que corresponde a 16,59%. Outras formas, como *INSTITUÍDO*, *DESCONTO EM FOLHA*, *CONVÊNIO COBRANÇA* e *CONVÊNIO PARCERIA*, representam percentuais menores, com 7,88%, 2,88%, 0,60% e 0,00%, respectivamente.

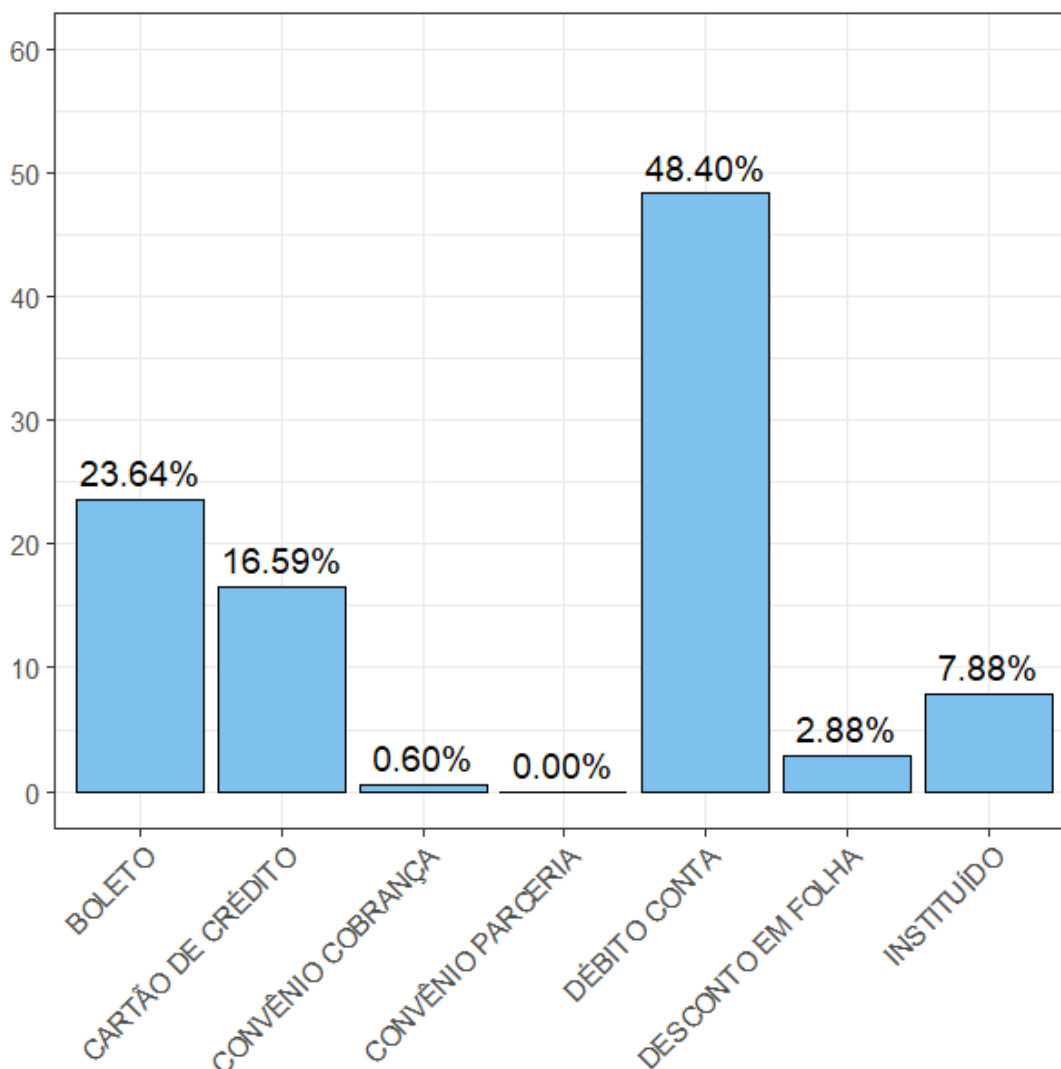


Figura 26: Desfecho Saldamento: Proporções por forma de cobrança (%)



5.1.6 Proporções por tipo de cobertura

O gráfico apresentado na Figura 27 ilustra a distribuição percentual do tipo de cobertura. Observa-se que a cobertura por *MORTE* é a mais prevalente, representando 93,24% do total, enquanto a cobertura por *INVALIDEZ* representa 6,76%.

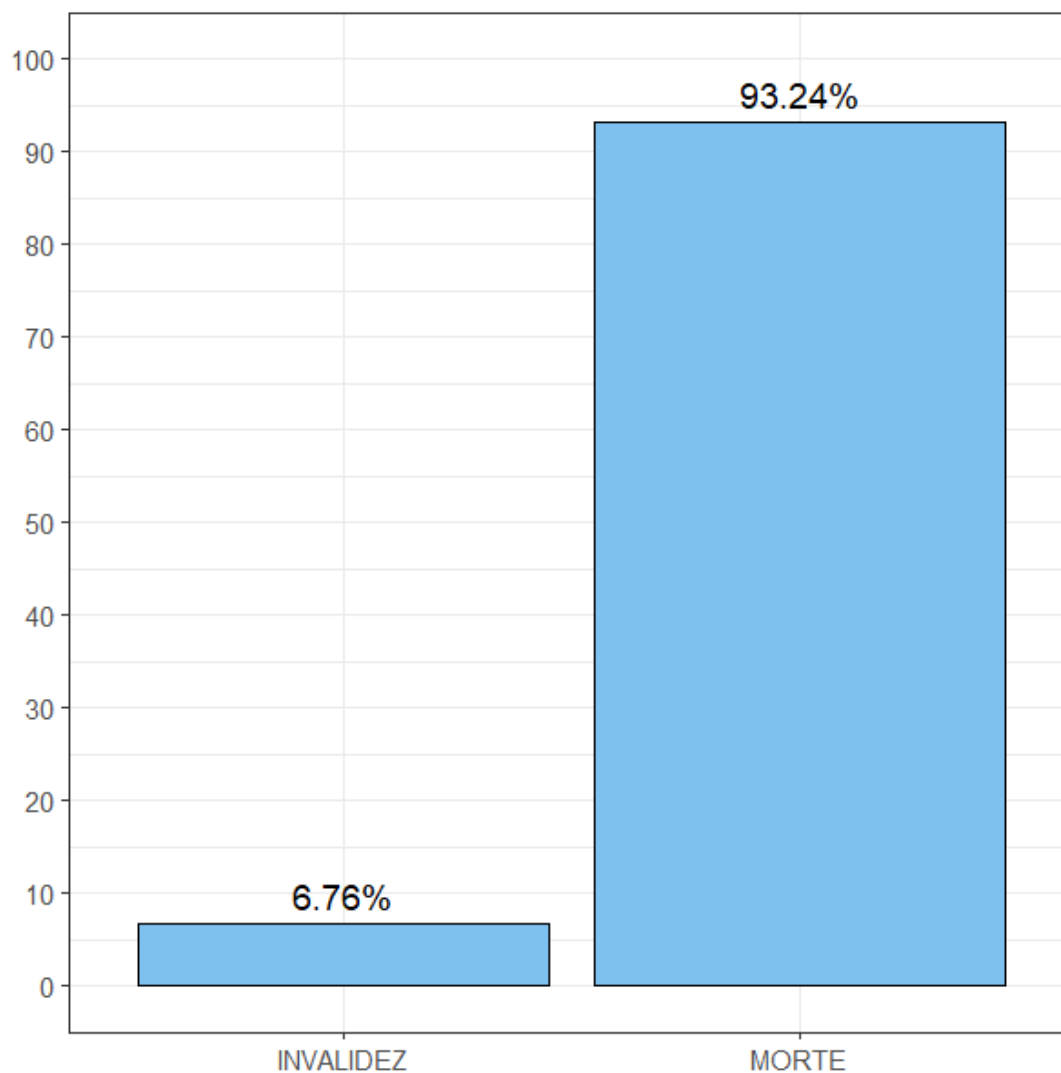


Figura 27: Desfecho Saldamento: Proporções por tipo de cobertura (%)



5.1.7 Clusterização

A clusterização foi realizada com o objetivo de simplificar a análise, preservando as informações relevantes. Note que as Figuras 21 – 23 apresentam o comportamento de cada covariável antes da clusterização, permitindo uma visualização clara das variabilidades.

Para compreender o comportamento das categorias associadas a cada covariável no desfecho saldamento, foi realizada a clusterização na covariável *Forma de Cobrança*, dada a sua complexidade e elevado número de categorias. As demais covariáveis também foram analisadas, mas a clusterização se mostrou particularmente necessária nesta covariável para facilitar a modelagem e interpretação dos dados.

Para determinar os clusters, foi utilizado o pacote `mclust` do software R, seguindo a metodologia apresentada na subseção 3.1. Uma descrição detalhada dos clusters formados para a covariável *Forma de Cobrança* está disponível no Apêndice A.

Com o objetivo de reduzir o número de categorias e facilitar a modelagem, foram definidos os seguintes clusters:

- Forma de Cobrança: De 7 categorias, foram definidos 3 clusters.

Durante a análise das combinações resultantes da clusterização da covariável *Forma de Cobrança* com outras covariáveis do estudo, identificamos 17 cenários distintos, embora o total esperado fosse 18. A ausência de informações no cenário específico sem categoriaCliente-FC2-INVALIDEZ resultou na inexistência desse cenário. A Figura 28 ilustra a função de risco observada para cada um desses cenários, destacando as diferenças e similaridades.

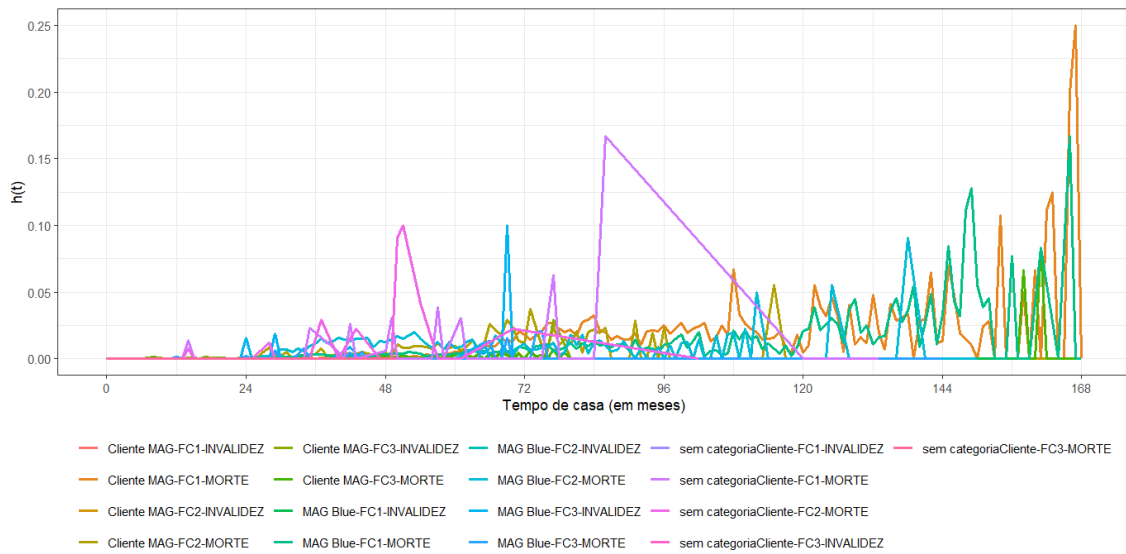


Figura 28: Desfecho Saldamento: Função de risco observada.

Para reduzir a complexidade e facilitar a análise, esses cenários foram reorganizados com base na similaridade das funções de risco, resultando em 6 cenários finais. Os agrupamentos são detalhados a seguir:

- CC4-FC4-TC4 agrupa os cenários:
Cliente MAG-FC3-INVALIDEZ e Cliente MAG-FC3-MORTE.
- CC5-FC5-TC5 agrupa os cenários:
Cliente MAG-FC1-INVALIDEZ; Cliente MAG-FC2-INVALIDEZ;
MAG Blue-FC1-INVALIDEZ; MAG Blue-FC2-INVALIDEZ;
MAG Blue-FC3-INVALIDEZ; MAG Blue-FC3-MORTE;
sem categoriaCliente-FC1-INVALIDEZ; sem categoriaCliente-FC1-MORTE;
sem categoriaCliente-FC2-MORTE; sem categoriaCliente-FC3-INVALIDEZ e
sem categoriaCliente-FC3-MORTE.

A Figura 29 apresenta a função de risco após o agrupamento, evidenciando a simplificação e a consolidação dos cenários.

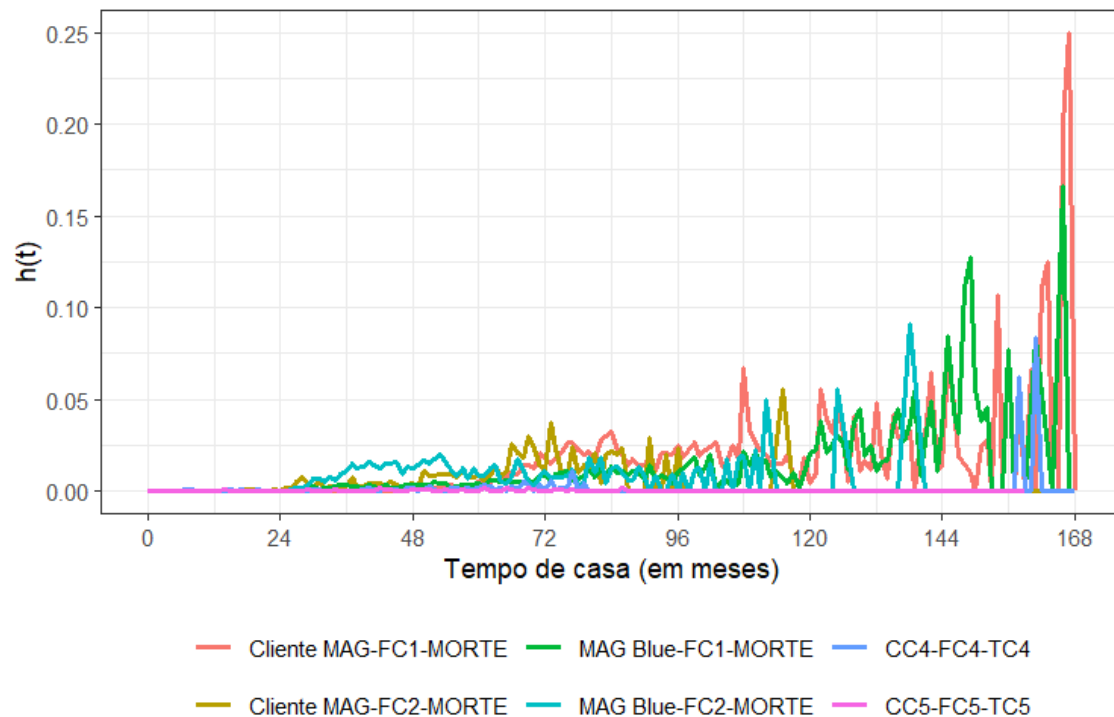


Figura 29: Desfecho Saldamento: Função de risco observada - após o agrupamento

Esses dois novos *clusters* oferecem uma base sólida para a análise, pois foram formados a partir da identificação de combinações de *clusters* que apresentavam funções de risco semelhantes. Ao agrupar as combinações com comportamentos semelhantes, o número total de *clusters* foi reduzido, o que simplifica a análise e a interpretação dos resultados.



5.2 Aplicação do Modelo de Mistura

Considerando as covariáveis *Categoria do Cliente*, *Forma de Cobrança* e *Tipo de Cobertura*, uma nova abordagem foi adotada para construir um modelo com o objetivo de aprimorar a explicação do tempo de persistência dos clientes. Os resultados são apresentados a seguir.

5.2.1 Cenários de Interesse Estudados

Para ajustar as curvas de persistência utilizando modelos de mistura, foram analisados diversos cenários de interesse, especificamente 6 cenários para o desfecho saldamento. A configuração desses cenários é detalhada a seguir. Para uma descrição detalhada dos clusters associados a covariável *Forma de Cobrança*, consulte o apêndice [A](#).

- **Cenário 1:**

Cliente MAG (CC) + Cluster 1 (FC) + MORTE (TC).

- **Cenário 2:**

Cliente MAG (CC) + Cluster 2 (FC) + MORTE (TC).

- **Cenário 3:**

MAG Blue (CC) + Cluster 1 (FC) + MORTE (TC).

- **Cenário 4:**

MAG Blue (CC) + Cluster 2 (FC) + MORTE (TC).

- **Cenário 5:**

Cliente MAG (CC) + Cluster 3 (FC) + INVALIDEZ (TC) e

Cliente MAG (CC) + Cluster 3 (FC) + MORTE (TC).

- **Cenário 6:**

Outros.



5.2.2 Representação dos Cenários nas Bases

A Tabela 7 contém os números de observações de cada um dos cenários apresentados na subseção 5.2.1 na base de dados.

Tabela 7: Desfecho Saldamento: Número de observações dos Cenários na base.

Cenário	observações	Frequência (%)
CENÁRIO 1	86.086	47,96
CENÁRIO 2	9.098	5,07
CENÁRIO 3	38.313	21,35
CENÁRIO 4	19.850	11,06
CENÁRIO 5	18.291	10,19
CENÁRIO 6	7.848	4,37



5.2.3 Parâmetros Estimados

As Tabelas 8 e 9 apresentam as estimativas pontuais e intervalares obtidas para o modelo de sobrevivência com mistura de Log-Normal.

Tabela 8: Desfecho Saldamento: Modelo de sobrevivência com mistura Log-Normal.

Parâmetro	Estimativa Pontual	IC (95%)
Peso da Mistura Componente 1 (η_1)	0,455	(0,435; 0,484)
Peso da Mistura Componente 2 (η_2)	0,324	(0,301; 0,351)
Peso da Mistura Componente 3 (η_3)	0,182	(0,173; 0,206)
Peso da Mistura Componente 4 (η_4)	0,036	(0,032; 0,039)
Precisão Componente 1 (ϕ_1)	9,909	(9,251; 10,690)
Precisão Componente 2 (ϕ_2)	55,155	(47,812; 61,012)
Precisão Componente 3 (ϕ_3)	0,875	(0,733; 0,989)
Precisão Componente 4 (ϕ_4)	18,968	(16,708; 21,602)
Intercepto Componente 1 (β_1)	4,687	(4,652; 4,711)
Intercepto Componente 2 (β_2)	4,363	(4,355; 4,374)
Intercepto Componente 3 (β_3)	5,532	(5,398; 5,815)
Intercepto Componente 4 (β_4)	4,662	(4,515; 4,826)
Coefficiente MAG Blue (CC) Componente 1 ($\beta_{MAGBlue,1}$)	-0,189	(-0,231; -0,159)
Coefficiente CC4 Componente 1 ($\beta_{CC4,1}$)	-0,588	(-49,123; 49,496)
Coefficiente CC5 Componente 1 ($\beta_{CC5,1}$)	5,916	(-43,344; 57,747)
Coefficiente FC2 Componente 1 ($\beta_{FC2,1}$)	-0,492	(-0,518; -0,467)
Coefficiente FC4 Componente 1 ($\beta_{FC4,1}$)	0,908	(-45,706; 51,439)
Coefficiente FC5 Componente 1 ($\beta_{FC5,1}$)	5,843	(-43,707; 56,898)
Coefficiente TC4 Componente 1 ($\beta_{TC4,1}$)	-0,422	(-50,151; 49,951)
Coefficiente TC5 Componente 1 ($\beta_{TC5,1}$)	5,771	(-44,946; 56,116)
Coefficiente MAG Blue (CC) Componente 2 ($\beta_{MAGBlue,2}$)	0,546	(0,520; 0,567)
Coefficiente CC4 Componente 2 ($\beta_{CC4,2}$)	1,857	(-48,873; 48,526)
Coefficiente CC5 Componente 2 ($\beta_{CC5,2}$)	6,869	(-42,663; 57,682)
Coefficiente FC2 Componente 2 ($\beta_{FC2,2}$)	0,591	(0,553; 0,677)
Coefficiente FC4 Componente 2 ($\beta_{FC4,2}$)	0,352	(-50,925; 51,246)
Coefficiente FC5 Componente 2 ($\beta_{FC5,2}$)	6,202	(-45,524; 56,871)
Coefficiente TC4 Componente 2 ($\beta_{TC4,2}$)	-1,468	(-49,832; 52,498)
Coefficiente TC5 Componente 2 ($\beta_{TC5,2}$)	5,870	(-43,143; 55,118)



Tabela 9: Desfecho Saldamento: Modelo de sobrevivência com mistura Log-Normal

Parâmetro	Estimativa Pontual	IC (95%)
Coeficiente MAG Blue (CC) Componente 3 ($\beta_{MAGBlue,3}$)	0,645	(0,429; 0,867)
Coeficiente CC4 Componente 3 ($\beta_{CC4,3}$)	-0,920	(-48,676; 52,112)
Coeficiente CC5 Componente 3 ($\beta_{CC5,3}$)	6,888	(-42,978; 55,846)
Coeficiente FC2 Componente 3 ($\beta_{FC2,3}$)	-0,489	(-0,655; -0,325)
Coeficiente FC4 Componente 3 ($\beta_{FC4,3}$)	0,876	(-50,356; 49,561)
Coeficiente FC5 Componente 3 ($\beta_{FC5,3}$)	6,314	(-42,445; 58,106)
Coeficiente TC4 Componente 3 ($\beta_{TC4,3}$)	0,506	(-50,989; 49,245)
Coeficiente TC5 Componente 3 ($\beta_{TC5,3}$)	6,062	(-40,268; 57,037)
Coeficiente MAG Blue (CC) Componente 4 ($\beta_{MAGBlue,4}$)	-1,000	(-1,176; -0,859)
Coeficiente CC4 Componente 4 ($\beta_{CC4,4}$)	-0,175	(-49,617; 48,317)
Coeficiente CC5 Componente 4 ($\beta_{CC5,4}$)	-0,923	(-51,788; 52,253)
Coeficiente FC2 Componente 4 ($\beta_{FC2,4}$)	0,053	(-0,012; 0,165)
Coeficiente FC4 Componente 4 ($\beta_{FC4,4}$)	0,532	(-51,325; 48,096)
Coeficiente FC5 Componente 4 ($\beta_{FC5,4}$)	-0,129	(-51,391; 51,127)
Coeficiente TC4 Componente 4 ($\beta_{TC4,4}$)	-0,040	(-49,036; 50,786)
Coeficiente TC5 Componente 4 ($\beta_{TC5,4}$)	0,949	(-50,453; 49,064)



5.2.4 Curvas de Persistência e Taxas de Falha

Nesta subseção, apresentamos gráficos comparando as curvas de persistência empíricas estimadas via Kaplan-Meier (KM) (linha sólida preta) com as ajustadas pelo modelo de sobrevivência com mistura de Log-Normal (MSMLN) (linha sólida azul). Ambas as curvas foram construídas com base nos dados completos a partir de 2010, permitindo avaliar a adequação do modelo ao comportamento de persistência observado. A seleção da curva final foi realizada com base no cálculo de métricas, resultando na escolha de um modelo com quatro componentes, que apresentou o melhor ajuste.

Além das curvas de persistência, também analisamos as taxas de falha, que fornecem uma medida do risco instantâneo de saldamento por parte do cliente. Em determinado tempo de interesse, a taxa de falha indica o risco do cliente saldar a inscrição naquele exato momento.

Assim, apresentaremos, nas Figuras 30 a 35, as curvas de persistência e as curvas de taxas de falha empíricas e ajustadas pelo MSMLN para cada um dos cenários estudados.

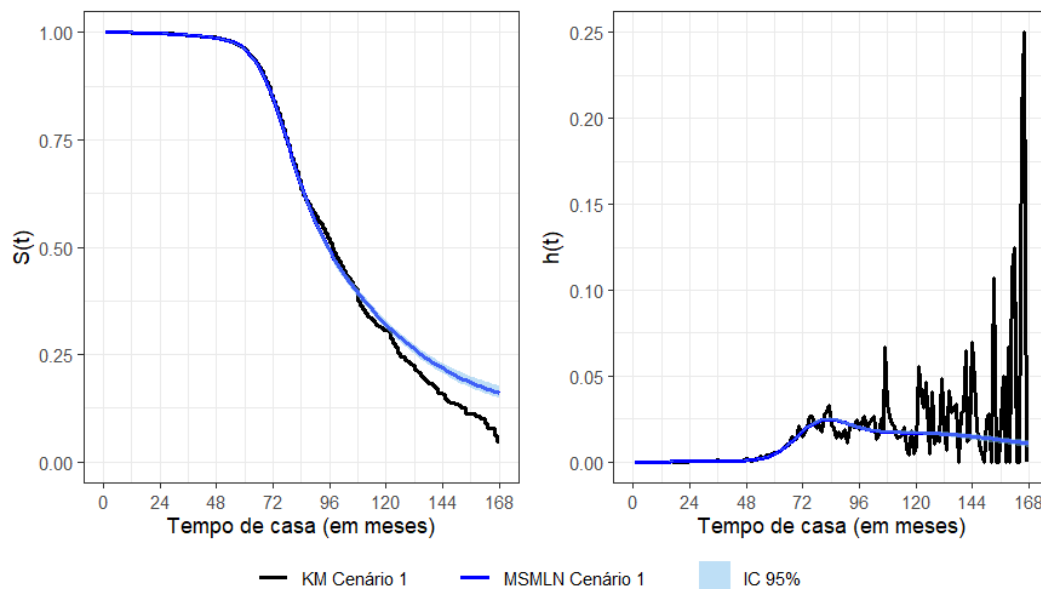


Figura 30: Desfecho Saldamento: Cenário 1 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica. Kaplan-Meier (linha sólida em preto) e MSMLN (linha sólida em azul).

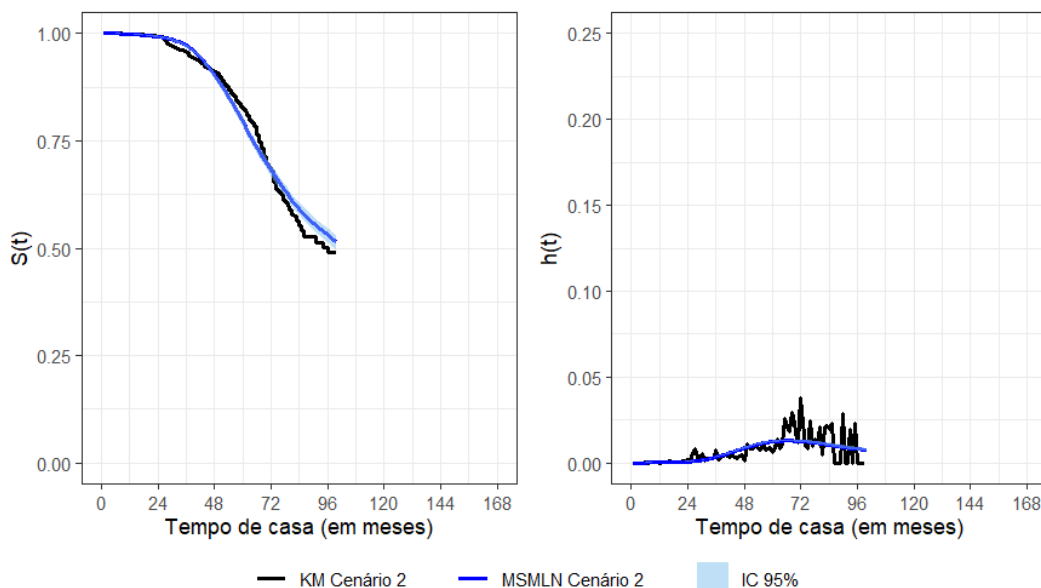


Figura 31: Desfecho Saldamento: Cenário 2 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica de Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

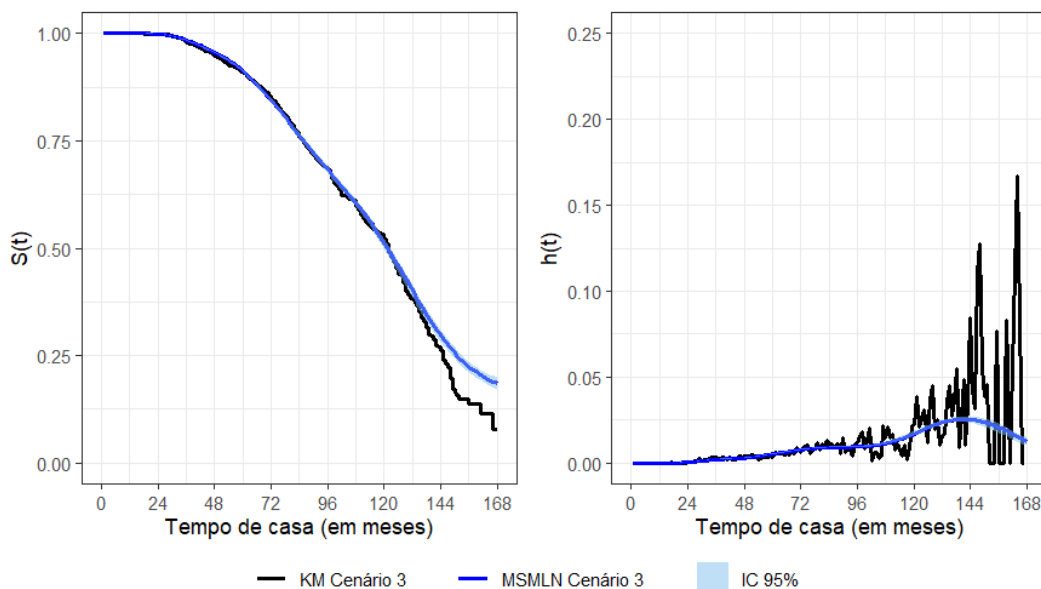


Figura 32: Desfecho Saldamento: Cenário 3 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica de Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

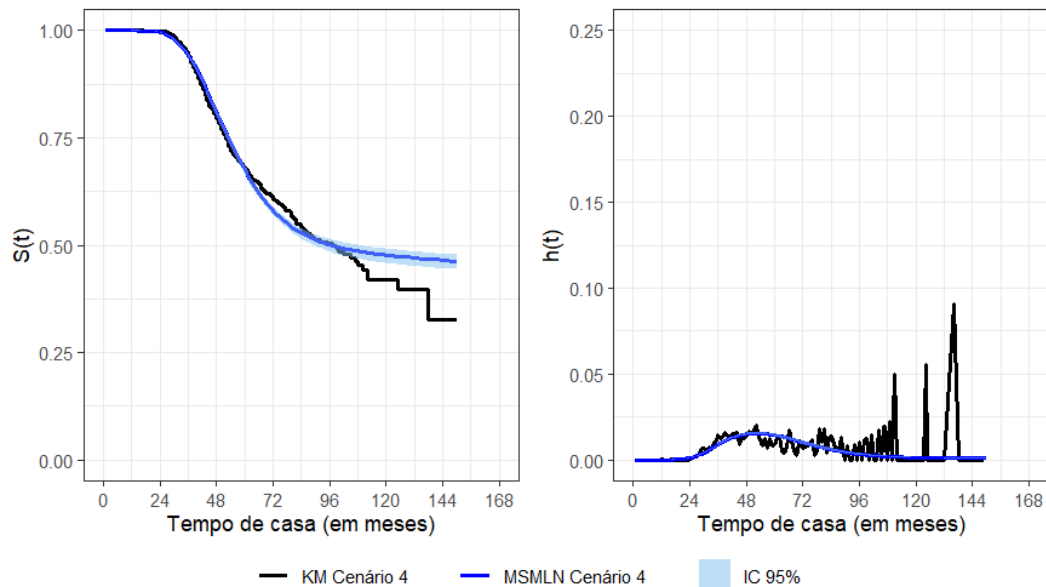


Figura 33: Desfecho Saldamento: Cenário 4 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica de Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

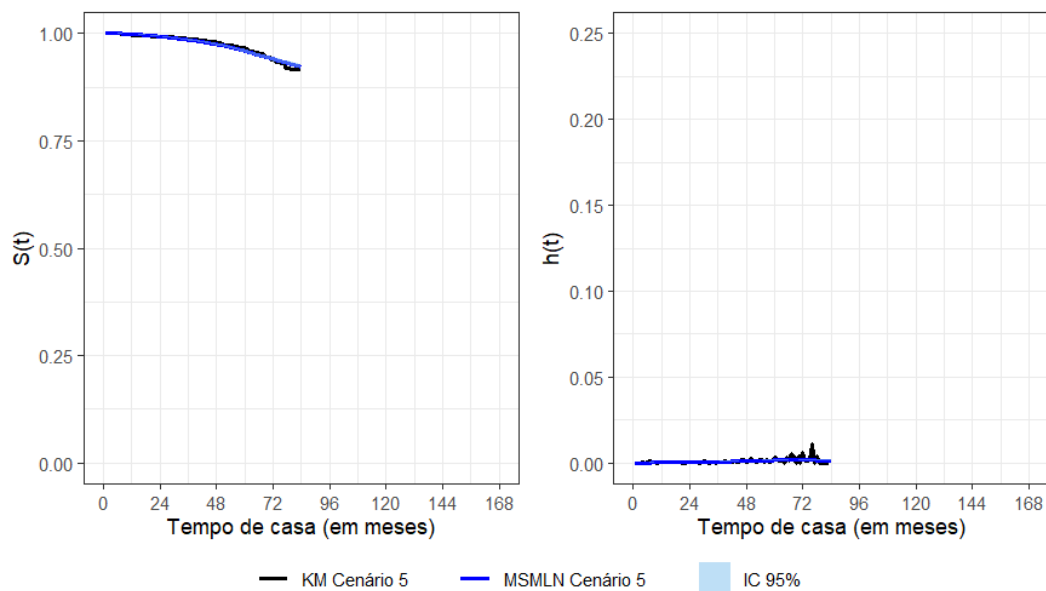


Figura 34: Desfecho Saldamento: Cenário 5 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica de Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

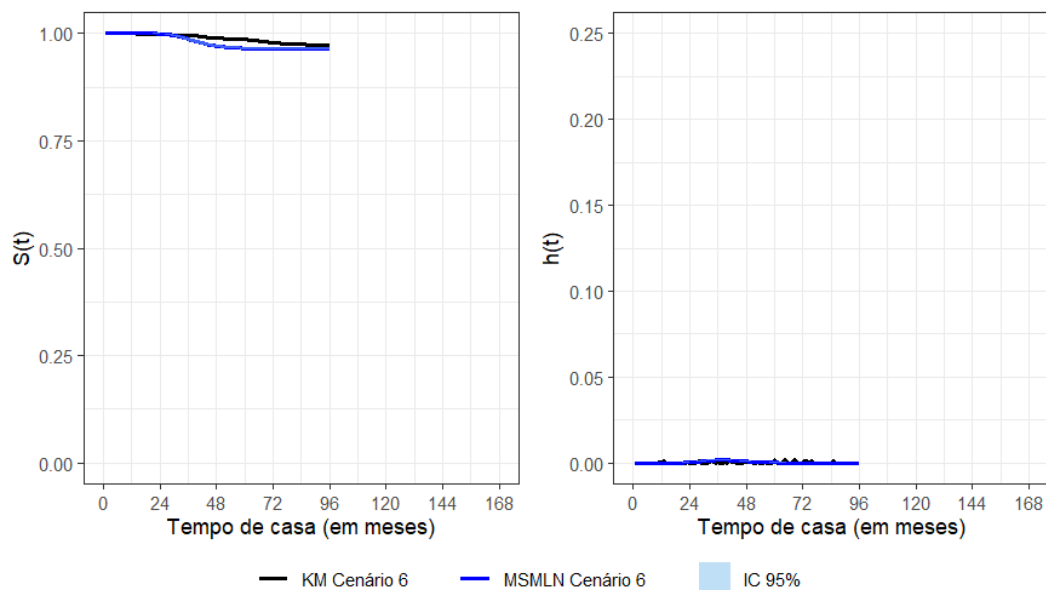


Figura 35: Desfecho Saldamento: Cenário 6 - Comparação das curvas de persistência (esquerda) e taxas de falha (direita) empírica de Kaplan-Meier (linha sólida em preto) e MSMLN ajustada (linha sólida em azul).

Observa-se, pelas Figuras 30 a 35, que o modelo de sobrevivência com mistura de Log-Normal (MSMLN) apresenta um bom ajuste às curvas empíricas em todos os cenários.



5.2.5 Comparações das Curvas de Persistência e Taxas de Falha para os cenários

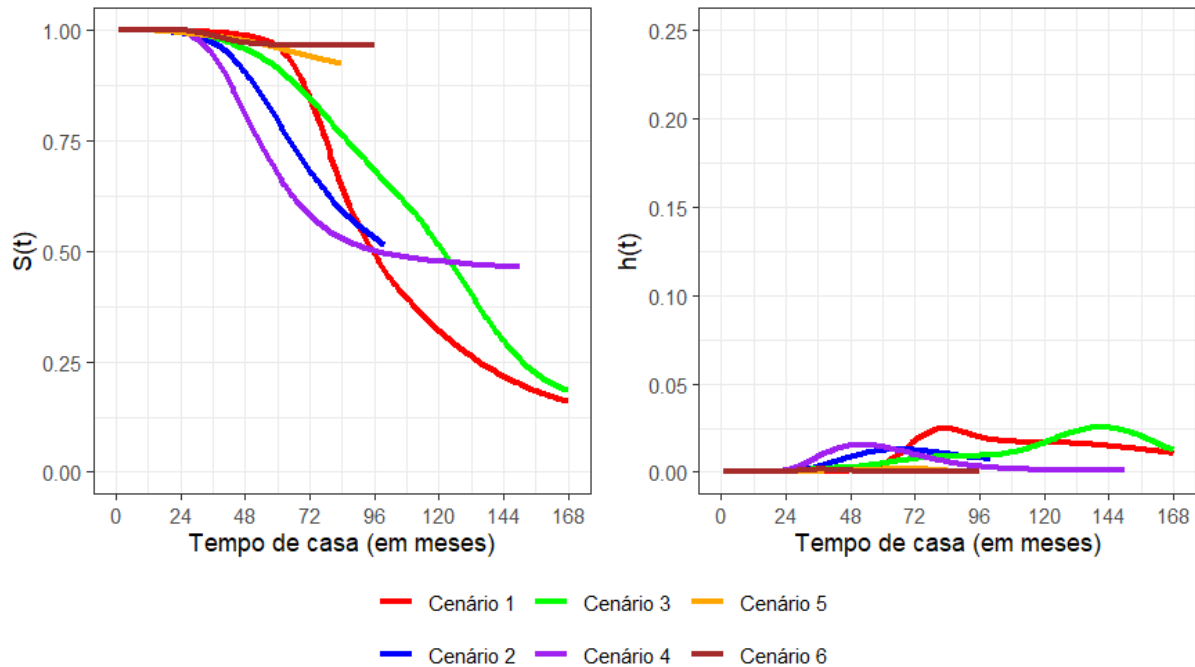


Figura 36: Desfecho Saldamento: Curvas de persistência ajustadas pelo MSMLN para todos os cenários.

Serão feitas na presente subseção, a partir das curvas apresentadas na Figura 36, comparações entre as persistências e taxas de falha observadas para diferentes cenários. Com isso, conseguimos comparar os riscos de saldamento instantâneos para clientes de diferentes cenários de interesse.



Cenário 1 versus Cenário 2

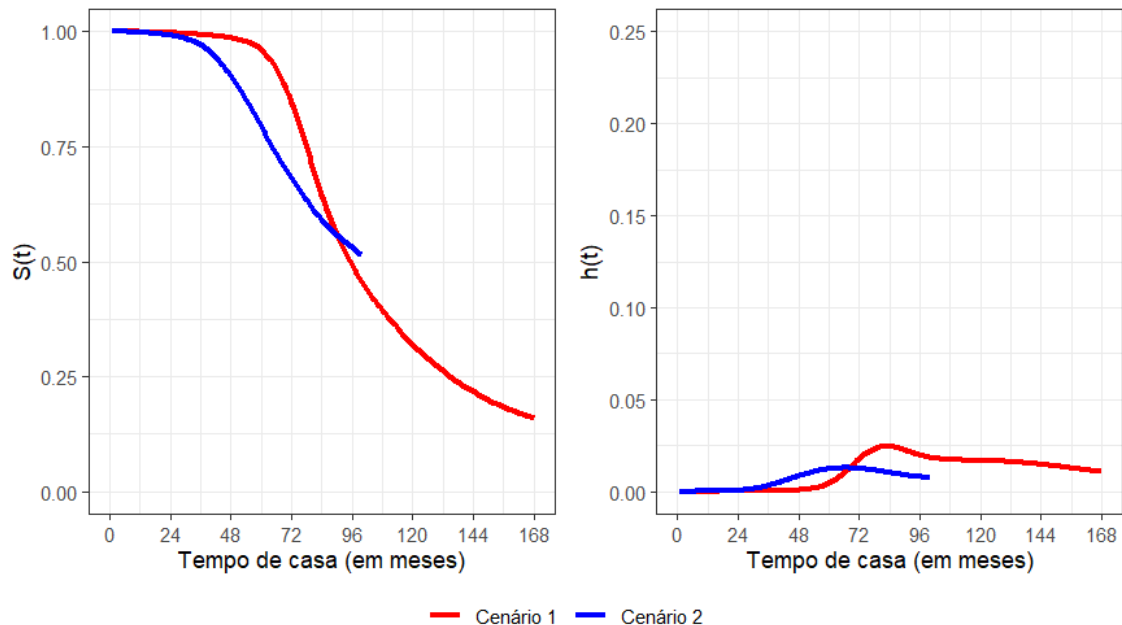


Figura 37: Desfecho Saldamento: Curvas de persistência e taxas de falha ajustadas pelo MSMLN para os Cenários 1 e 2

Conforme descrição apresentada na subseção 5.2.1, é possível notar que a diferença observada entre os Cenários 1 e 2 se dá apenas pela variável *Forma de Cobrança*. Enquanto no Cenário 1 os clientes são do Cluster 1 dessa variável (*BOLETO*, *CONVÊNIO COBRANÇA*, *CONVÊNIO PARCEIRA* e *DÉBITO CONTA*), no Cenário 2 os clientes são do Cluster 2 (*CARTÃO DE CRÉDITO*). A partir disso, é interessante comparar as curvas de persistência de ambos os cenários para verificar qual o impacto que essa variável em especial pode gerar no tempo de permanência do cliente.

Como mostra a Figura 37, os clientes do Cenário 1 tendem a persistir mais do que clientes do Cenário 2. Dessa forma, podemos concluir que, para *Categoria do Cliente* Cliente MAG e *Tipo de Cobertura* Morte, aqueles que pertencem ao Cluster 1 de *Forma de Cobrança* tendem a permanecer na empresa por mais tempo do que aqueles que pertencem ao Cluster 2. Ao observar a taxa de falha, verifica-se que, ao longo de quase todo o período observado, o risco de saldamento dos clientes do Cenário 1 é menor do que o dos clientes do Cenário 2. No entanto, há um aumento no risco em torno dos 72



meses para os clientes do Cenário 1, momento em que o risco se iguala ao do Cenário 2 e, a partir desse ponto, o risco do Cenário 1 passa a ser maior.

Se fizermos as comparações Cenário 3 *versus* Cenário 4, onde a única diferença também se dá apenas pela variável *Forma de Cobrança*, chegaremos às mesmas conclusões, de que clientes do Cluster 1 dessa variável persistem mais do que clientes do Cluster 2.

Sendo assim, visando a captação de novos clientes que persistam por mais tempo na empresa, pode ser interessante priorizar a forma de cobrança do Cluster 1 em relação aos do Cluster 2, visto que, de forma geral, esse Cluster vai apresentar melhores resultados sob as condições contratuais inicialmente estabelecidas.

Cenário 1 *versus* Cenário 3

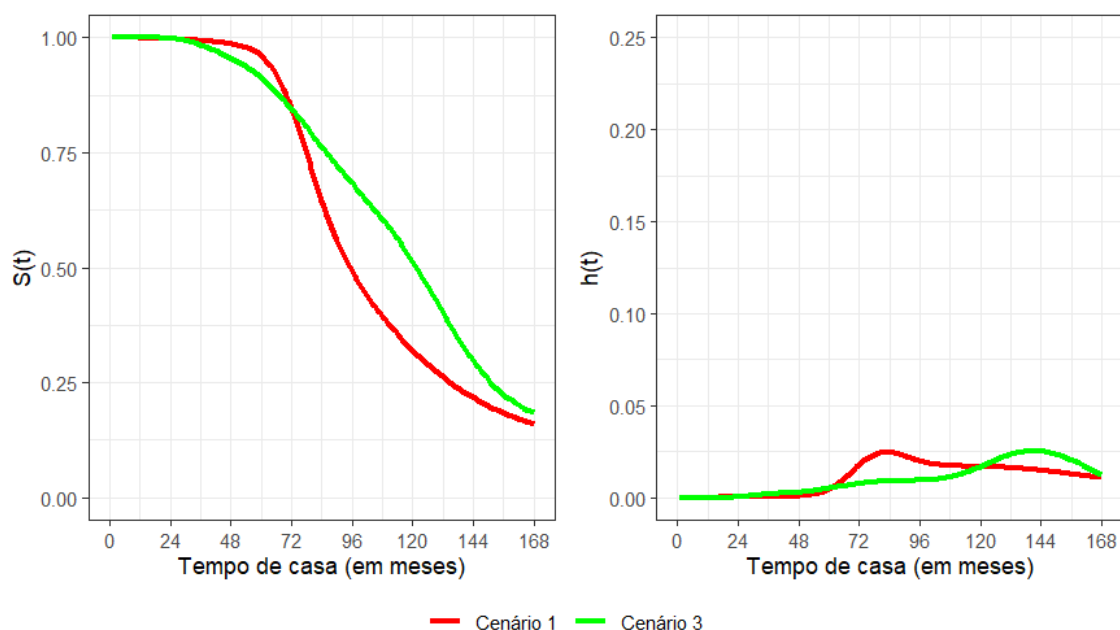


Figura 38: Desfecho Saldamento: Curvas de persistência e taxas de falha ajustadas pelo MSMLN para os Cenários 1 e 3

Conforme descrição apresentada na subseção 5.2.1, é possível notar que a diferença observada entre os Cenários 1 e 3 se dá apenas pela variável *Categoria do Cliente*. Enquanto no Cenário 1 os clientes são *Cliente MAG*, no Cenário 3 os clientes são *MAG Blue*. A partir disso, é interessante comparar as curvas de persistência de ambos os



cenários para verificar qual o impacto que essa variável em especial pode gerar no tempo de permanência do cliente.

Como mostra a Figura 38, ao longo dos primeiros 24 meses as curvas praticamente se sobrepõem, indicando uma persistência semelhante entre os dois cenários. Entre os tempos 24 e 72 meses, os clientes do Cenário 1 apresentam uma persistência maior em comparação aos do Cenário 3. Após 72 meses, entretanto, os clientes do Cenário 3 persistem mais do que os do Cenário 1.

Dessa forma, podemos concluir que, para clientes do Cluster 1 de *Forma de Cobrança e Tipo de Cobertura* Morte, aqueles que pertencem ao *Categoria do Cliente* MAG Blue tendem a permanecer na empresa por mais tempo do que aqueles que pertencem ao Cliente MAG a partir do tempo 72. Ao observar a taxa de falha, verifica-se que, na maior parte do tempo observado, o risco de saldamento dos clientes do Cenário 3 é maior do que o dos clientes do Cenário 1, com exceção de um aumento na taxa de falha por volta dos 72 meses para o Cenário 1.

Sendo assim, visando a captação de novos clientes que persistam por mais tempo na empresa, pode ser interessante priorizar os clientes da categoria MAG Blue em relação aos Cliente MAG, visto que, de uma forma geral, esse cluster vai apresentar melhores resultados sob as condições contratuais inicialmente estabelecidas.



5.3 Probabilidades de Persistência

Na presente subseção, serão calculadas algumas probabilidades de persistência de interesse para o desfecho saldamento. A partir de informações fornecidas para um cliente, calcularemos com que probabilidade este cliente irá persistir por um certo tempo na empresa.

5.3.1 Exemplo 1

Um cliente novo na empresa é classificado como um Cliente MAG, com forma de cobrança Débito Conta e tipo de cobertura Invalidez. Qual a probabilidade desse cliente permanecer na empresa por 4 anos?

De acordo com o apêndice A, ao final do relatório, onde encontra-se uma descrição dos Clusters para todas as variáveis estudadas, podemos concluir que o cliente em questão pertence ao Cluster 1 de *Forma de Cobrança*. Pela descrição dos cenários em 5.2.1, podemos verificar que o cliente em questão está inserido no que chamamos de Cenário 1.

Utilizando-se do modelo ajustado, encontramos uma estimativa $\widehat{S}(t)$ para a função de sobrevivência no tempo de interesse, no caso $t = 48$ meses. Dessa forma, encontramos que $\widehat{S}(48) = 0,9868$, e portanto a probabilidade do cliente em questão permanecer na empresa por 4 anos é de aproximadamente 98,68%.

5.3.2 Exemplo 2

Suponha agora que passaram-se 2 anos e que o mesmo cliente do Exemplo 1 ainda permanece na empresa. Queremos atualizar a probabilidade de esse cliente permanecer 4 anos na empresa, dado que ele já permaneceu por 2 anos. Para isso, faremos:

$$P(T > 48 \mid T > 24) = \frac{P((T > 48) \cap (T > 24))}{P(T > 24)} = \frac{P(T > 48)}{P(T > 24)} = \frac{S(48)}{S(24)}. \quad (18)$$

Já calculamos no Exemplo 1 o valor de $\widehat{S}(48)$. Precisamos agora calcular $\widehat{S}(24)$, de forma análoga ao que foi feito para calcular $\widehat{S}(48)$, porém com $t = 24$ meses. Ao fazer os cálculos, obtemos que $\widehat{S}(24) = 0,9975$. Dessa forma, temos que:



$$P(T > 48 \mid T > 24) = \frac{S(48)}{S(24)} = \frac{0,9868}{0,9975} = 0,9893. \quad (19)$$

Isto é, após permanecer por 2 anos, a probabilidade do cliente em questão persistir na empresa por 4 anos aumentou de 98,68% para 98,93%. Esse aumento é coerente, uma vez que, por já haver permanecido por algum tempo como cliente da empresa, é mais provável que ele seja um cliente “fiel”. Além disso, agora ele precisa permanecer por menos tempo na empresa para completar 4 anos na mesma, isto é, ele só precisa permanecer mais 24 meses, e não mais 48 meses, uma vez que 24 meses já se passaram.

5.3.3 Exemplo 3

Um novo cliente da empresa possui exatamente as mesmas características do cliente estudado nos dois exemplos anteriores, porém este novo cliente é classificado como Cartão de Crédito, ao contrário do cliente anterior, que era classificado como Débito em Conta. Desejamos saber com que probabilidade esse novo cliente permanecerá na empresa pelo período de 4 anos.

Pelas informações fornecidas, temos que este novo cliente pertence a Cliente MAG de *Categoria do Cliente*, ao Cluster 2 de *Forma de Cobertura* e ao *Tipo de Cobertura* Morte. Pela descrição dos cenários em 5.2.1, podemos verificar que o cliente em questão está inserido no que chamamos de Cenário 2.

Mais uma vez utilizando-se do modelo ajustado, obtemos que $\widehat{S}(48) = 0,9047$. Ou seja, a probabilidade do cliente em questão permanecer na empresa por 4 anos é de aproximadamente 90,47%, e portanto menor do que a probabilidade do cliente do Exemplo 1 permanecer pelo mesmo tempo (98,68%). Tal resultado é coerente, uma vez que, conforme verificado na subseção 5.2.5 e ilustrado na Figura 37, os clientes que utilizam Débito em Conta, em geral, apresentam maior persistência na empresa do que aqueles que optam pelo Cartão de Crédito.



6 Considerações Finais

Neste relatório, foram analisados dois desfechos distintos de produtos de capitalização: resgate e saldamento. Para cada desfecho, foi aplicada uma filtragem específica, separando os casos de resgate e saldamento, de forma a ajustar os dados à análise correspondente. A modelagem das curvas de persistência considerou a abordagem de modelos Bayesiano de sobrevivência com mistura Log-Normal (MSMLN), implementada por meio do pacote `lrmixsurv` no software R, proporcionando uma avaliação robusta das características dos cenários em relação à sua persistência e risco. O modelo é capaz de capturar mudanças no padrão de risco além de acomodar picos de risco ao longo do tempo.

É importante ressaltar, novamente, que o desfecho saldamento não possui o *Tipo de Cobertura* Sobrevivência, ao contrário do desfecho resgate, evidenciando diferenças importantes entre os tipos de cobertura analisados. A modelagem foi realizada separadamente para cada desfecho, considerando as especificidades de cada um. Dessa forma, *clusters* distintos foram identificados nas categorias de covariáveis, cada um apresentando curvas de sobrevivência e perfis de risco específicos para os tempos até a ocorrência do desfecho (resgate ou saldamento). Além disso, cálculo da probabilidade de persistência foi considerado, de forma a tentar compreender o tempo de persistência em ambos os desfechos para diferentes tipos de cliente da empresa.



7 Equipe Responsável

Coordenador Mário M. C. de Oliveira (PhD) – UFRJ

Prof. Thaís C. O. Fonseca (PhD) - UFRJ

Prof. Viviana G. R. Lobo (D.Sc.) - UFRJ

Prof. Mariane B. Alves (D.Sc.) - UFRJ

Prof. Bruno Costa (PhD) – UFRJ

Prof. Rafael B. de R. Borges (D.Sc.) – UERJ

Prof. Milton Ramirez (D.Sc.) – UFRJ

Mário M. C. de Oliveira
Coordenador do LabMA/UFRJ



A Descrição completa das covariáveis e dos clusters

A.1 Desfecho Resgate: Conjunto de covariáveis

- **Categoria do Cliente:** Cliente MAG, MAG Blue e sem categoriaCliente;
- **Forma de Cobrança:** BOLETO, CARTÃO DE CRÉDITO, CONVÊNIO COBRANÇA, CONVÊNIO PARCERIA, DÉBITO CONTA, DESCONTO EM FOLHA e INSTITUÍDO;
- **Tipo de Cobertura:** INVALIDEZ, MORTE e SOBREVIVÊNCIA.

A.2 Desfecho Resgate: Descrição dos clusters por covariável

- **Categoria do Cliente**
 - **Cluster 1:** Cliente MAG e sem categoriaCliente;
 - **Cluster 2:** MAG Blue.
- **Forma de cobrança**
 - **Cluster 1:** BOLETO, CONVÊNIO COBRANÇA, DÉBITO CONTA;
 - **Cluster 2:** CARTÃO DE CRÉDITO e CONVÊNIO PARCERIA;
 - **Cluster 3:** DESCONTO EM FOLHA;
 - **Cluster 4:** INSTITUÍDO.
- **Tipo de cobertura**
 - **Cluster 1:** INVALIDEZ e MORTE;



- **Cluster 2:** SOBREVIVÊNCIA.

A.3 Desfecho Saldamento: Conjunto de covariáveis

- **Categoria do Cliente:** Cliente MAG, MAG Blue e sem categoriaCliente;
- **Forma de Cobrança:** BOLETO, CARTÃO DE CRÉDITO, CONVÊNIO COBRANÇA, CONVÊNIO PARCERIA, DÉBITO CONTA, DESCONTO EM FOLHA e INSTITUÍDO;
- **Tipo de Cobertura:** INVALIDEZ e MORTE.

A.4 Desfecho Saldamento: Descrição dos clusters por covariável

- **Forma de cobrança**
 - **Cluster 1:** BOLETO, CONVÊNIO COBRANÇA, CONVÊNIO PARCERIA, DÉBITO CONTA;
 - **Cluster 2:** CARTÃO DE CRÉDITO;
 - **Cluster 3:** DESCONTO EM FOLHA, INSTITUÍDO.



B Descrição da Censura

No presente estudo, os *status item contratado* foram utilizados como base para a criação de uma nova coluna denominada *status*, que possui as seguintes categorias: *censura*, *resgate* e *saldamento*. A categoria *censura* abrange situações em que não foi observado o evento de interesse (resgate ou saldamento). Abaixo estão descritos os *status item contratado* classificados como *censura*:

- PECÚLIO - B03;
- INVALIDEZ (PAGAMENTO ÚNICO IPD) - B09;
- AUXÍLIO FUNERAL - B07;
- LIQUIDAÇÃO ESPECIAL BENEFÍCIO - B05;
- PRESCRIÇÃO BENEFÍCIO - C11;
- CANCELADO POR DEPENDÊNCIA - C21;
- ÓBITO - C03;
- SOBREVIVÊNCIA - B08;
- PRAZO CUMPRIDO - C06;
- ATIVA 1 ATRASO - A01;
- ATIVA 2 ATRASOS - A02;
- ATIVA 3 ATRASOS - A03;
- ATIVA 4 ATRASOS - A04;
- ATIVA 5 ATRASOS - A05;
- ATIVA EM DIA - A00;
- REMIDO - D02.



Referências

- Enrico Antonio Colosimo and Suely Ruiz Giolo. Análise de sobrevivência aplicada. Editora Blucher, 2006.
- Sylvia Frühwirth-Schnatter. Finite mixture and Markov switching models. Springer Science & Business Media, 2006.
- D. Gamerman and H.F. Lopes. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Texts in Statistical Science. Taylor & Francis, 2 edition, 2006.
- Viviana G.R. Lobo, Thaís C.O. Fonseca, and Mariane B. Alves. Lapse risk modeling in insurance: a bayesian mixture approach. Annals of Actuarial Science, 18:126–151, 2024.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL <https://www.R-project.org/>.
- Martin A Tanner and Wing Hung Wong. The calculation of posterior distributions by data augmentation. Journal of the American statistical Association, 82(398):528–540, 1987.