



**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS**



Gabriel de Freitas Monteiro Pereira

**Análise de sobrevivência aplicada ao tempo de inadimplência de clientes de uma
empresa de crédito**

São Cristóvão - SE

2022

Gabriel de Freitas Monteiro Pereira

Análise de sobrevivência aplicada ao tempo de inadimplência de clientes de uma empresa de crédito

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como parte dos requisitos para obtenção do grau de Bacharel em Ciências Atuariais.

Orientador: Prof. Dr. Cleber Martins Xavier

São Cristóvão - SE

2022

*“Se fracassar, ao menos que fracasse ousando grandes feitos,
de modo que a sua postura não seja nunca a dessas almas frias e
tímidas que não conhecem nem a vitória nem a derrota
(Weak to be, unless daring great deeds,
so that your posture is never to those cold souls and
shy that know neither victory nor defeat)”
(Theodore Roosevelt)*

Resumo

Com aumento do número de empresas que oferecem crédito no país e o grande volume de produtos ofertados, que atingem os mais diversos perfis de clientes, faz-se necessário o controle de risco efetivo por parte das instituições, visando garantir a melhora dos índices de inadimplência, fraude e cobrança, que são variáveis imprescindíveis para solvência de uma instituição financeira. Sendo assim, além de um modelo de *credit scoring* que leve em consideração a política de crédito e o apetite de risco da instituição, a análise dos clientes que já fazem parte do portfólio através dos modelos nomeados de *behavior scoring* vem ganhando cada vez mais espaço, por serem capazes de reavaliar o cliente e nortear as decisões de curto prazo da empresa. Entre as diversas técnicas utilizadas para construção desses modelos, a análise de sobrevivência vem ganhando espaço por ser capaz de analisar o tempo até evento de interesse. Nesse contexto, o objetivo desse trabalho é aplicar as técnicas da análise de sobrevivência na construção de um modelo comportamental, para 94.407 clientes de uma empresa multinacional varejista que oferece o produto cartão de crédito, entendendo a relação das variáveis selecionadas no processo de inadimplência. O modelo paramétrico selecionado apresentou ajuste satisfatório para todas as covariáveis selecionadas, permitindo a identificação de algumas características que influenciam o tempo até inadimplência dos clientes observados.

Palavras-chave: *Behavior Scoring*, Inadimplência, Modelo comportamental, controle de risco.

Abstract

With the increase in the number of companies offering credit in the country and the large volume of products offered, which reach the most diverse customer profiles, it is necessary for institutions to effectively control risk, in order to guarantee the improvement of credit indices. default, fraud and collection, which are essential variables for the solvency of a financial institution. Therefore, in addition to a *credit scoring* model that takes into account the institution's credit policy and risk appetite, the analysis of clients that are already part of the portfolio through named *behavior scoring models* has been gaining more and more space, as they are able to reassess the customer and guide the company's short-term decisions. Among the various techniques used to build these models, survival analysis has been gaining ground for being able to analyze the time until the event of interest. In this context, the objective of this work is to apply survival analysis techniques in the construction of a behavioral model, for 94,407 customers of a multinational retail company that offers the credit card product, understanding the relationship of the selected variables in the default process . The selected parametric model presented adjustment for all selected covariates, allowing the identification of some characteristics that influence the time until default of the observed customers.

Keywords: Behavior Scoring, Default, Behavioral model, risk control.

Lista de ilustrações

Figura 1 – Fluxo da concessão de crédito utilizando um modelo de <i>Credit Scoring</i>	15
Figura 2 – Perfil de renda dos clientes da empresa A.	26
Figura 3 – Perfil de idade na aquisição dos clientes da empresa A de 2000 a 2022 . . .	26
Figura 4 – Informação de UF, obtido na aquisição dos clientes da empresa A de 2000 a 2022	27
Figura 5 – Comparação entre perfil de idade na aquisição dos clientes cancelados/ativos e inadimplentes.	27
Figura 6 – Atividade dos clientes em relação a Faixa de risco da obtida na aquisição dos clientes entrantes de 2022.	29
Figura 7 – Classificação de risco aplicada no modelo comportamental existente da empresa.	30
Figura 8 – Perfil de idade dos clientes da empresa A	30
Figura 9 – Estimativa de Kaplan-Meier para a sobrevivência, em semanas, dos clientes que entraram em inadimplência.	31
Figura 10 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Faixa de risco na aquisição. . . .	32
Figura 11 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Idade.	33
Figura 12 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Renda.	33
Figura 13 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Unidade Federativa.	34
Figura 14 – Gráficos das Curvas de sobrevivência estimadas por Kaplan-Meier <i>versus</i> as sobrevivências estimadas pelos modelos Exponencial, de Weibull e log-normal.	35
Figura 15 – Gráficos t <i>versus</i> $-\log(\hat{S}(t))$, $\log(t)$ <i>versus</i> $\log(-\log(\hat{S}(t))) \Phi^{-1} \log(t)$. . .	35
Figura 16 – Curva de sobrevivência estimadas por Kaplan-Meier <i>versus</i> as curvas de sobrevivência estimadas pelos modelos Exponencial, de Weibull e log-normal em relação ao tempo, com base no evento de inadimplência	36
Figura 17 – Análise dos resíduos de Cox-Snell do modelo de regressão Weibull ajustado.	37
Figura 18 – Sobrevivência do grupo com a menor e maior sobrevida possível com base no modelo de regressão Weibull ajustado.	38

Lista de quadros

Quadro 1 – Estimativa dos parâmetros do modelo de regressão Weibull	38
---	----

Lista de tabelas

Tabela 1 – Frequência relativa entre a idade e renda dos clientes da empresa A de 2000 a 2022.	28
Tabela 2 – Definição da faixa de risco dos clientes com base no modelo de crédito aplicado na empresa A.	28

Lista de abreviaturas e siglas

SFN - Sistema Financeiro Nacional

LGPD - Lei Geral de Proteção de dados

Sumário

1	INTRODUÇÃO	10
2	OBJETIVOS	12
2.1	Geral	12
2.2	Específicos	12
3	JUSTIFICATIVA	13
4	REVISÃO LITERÁRIA	14
5	METODOLOGIA	18
5.1	Material	18
5.1.1	Base de dados	18
5.2	Recursos	18
5.3	Métodos	18
5.3.1	Seleção de Variáveis	18
5.3.2	Análise de sobrevivência	19
5.3.3	Estimador de Kaplan-Meier	21
5.3.4	Modelos Paramétricos	22
5.3.4.1	Distribuição Exponencial	23
5.3.4.2	Distribuição Weibull	23
5.3.4.3	Distribuição Log-Normal	24
5.3.4.4	Escolha do modelo e estimação dos parâmetros	24
6	RESULTADOS	26
6.1	Análise Descritiva e Exploratória	26
6.2	Estimador de Kaplan-Meier	31
6.3	Ajuste do modelo probabilístico	34
7	CONCLUSÃO	39
	REFERÊNCIAS	40
	APÊNDICE A – MODELOS	42

1 INTRODUÇÃO

O Sistema Financeiro Nacional (SFN) surge para integrar os agentes superavitários e deficitários, a fim de suprir a necessidade de consumo das empresas, famílias e governo através da transferência de recursos. [Presente \(2019\)](#) define o SFN da seguinte forma: Conjunto de instituições e instrumentos que viabilizam o fluxo financeiro entre os poupadores e os tomadores na economia.

Nesse contexto, o mercado de crédito pode ser entendido como uma subdivisão importante do SFN, tendo como principal função a intermediação de valores com prazo médio/longo através das Instituições financeiras, que são órgãos públicos ou privados com capacidade de assessorar, intermediar, aplicar e custodiar os recursos de terceiros ([ABREU, 2017](#)).

Desde a origem do mercado de crédito no Brasil, em meados de 1808, o ato de realizar uma troca de "recursos" esteve diretamente ligado ao desenvolvimento econômico, à evolução da moeda e do sistema jurídico, impactando de forma direta ou indireta a renda, a oferta, demanda e evolução de maneira geral do país ([MANOS, 2006](#)).

A velocidade e facilidade com que as interações entre os *players* ocorrem, aliado a um movimento disruptivo nas tecnologias empregadas no sistema financeiro e o aumento do número de *fintechs* no mercado são alguns dos fatores que impulsionam a maior disponibilidade de crédito no mercado e a constante evasão de clientes buscando melhores experiências. ([PRESENTE, 2019](#))

Em um mercado com tantas empresas oferecendo crédito através dos diversos produtos e atingindo tipos diferentes de público, o fator controle de risco sobre as variáveis inadimplência, prejuízo e fraude assume um estado primordial na solvência da instituição a longo prazo. Visando facilitar, ampliar e desenvolver um padrão nas decisões das empresas, os modelos de crédito tem ganhado cada vez mais espaço no mercado.

Podemos dividir o processo de concessão de crédito em dois grandes estágios. A primeira, conhecida como *front-end* é a decisão de oferecer crédito ao cliente através da análise de perfil prévia utilizando dados disponibilizados por centrais de informações de crédito, em grande parte se valendo de modelos que utilizam como base o processo de árvore de decisão ou regressão logística, retornando um score (*credit scoring*) que é utilizado para decidir a entrada ou não do cliente na empresa. O segundo é conhecido como modelo comportamental (*behavior scoring*) que visa através de dados do comportamento do cliente, prever a inadimplência e entender as características mais influentes nesse processo, com o objetivo de reter os clientes lucrativos e de baixo risco ([ARAÚJO, 2006](#)).

Não é comum a existência de uma regra geral que norteia a construção de um modelo comportamental único entre as empresas, podendo ser aplicadas técnicas como: redes neurais, regressão logística, algoritmos gerais e de forma mais recente a análise de sobrevivência, que

vem ganhando espaço por ser capaz de englobar a variável tempo e retornar uma previsão do momento até a inadimplência do cliente.

De modo geral, esse trabalho tem o objetivo de aplicar a análise de sobrevivência para construção de um modelo comportamental, utilizando-se de dados que englobam o perfil dos clientes e o resultado do modelo de crédito aplicado na aquisição, entendendo a influência das variáveis selecionadas na inadimplência do cliente.

2 OBJETIVOS

2.1 Geral

O presente trabalho tem como objetivo estudar o tempo até a inadimplência dentro de uma empresa multinacional varejista que oferece o produto de cartão de crédito, buscando encontrar uma relação com o perfil do cliente e integrar com o modelo comportamental existente na empresa.

2.2 Específicos

- Traçar o perfil geral dos clientes da empresa e comparar com o dos clientes inadimplentes;
- Aplicar as técnicas de análise de sobrevivência para entender a relação perfil *versus* inadimplência;
- Discutir os resultados e apresentar as principais conclusões acerca do modelo ajustado;

3 JUSTIFICATIVA

A solvência e o processo de controle do risco, são considerados pontos chaves para uma instituição financeira. O constante crescimento no número de instituições e na abrangência de produtos que tenham como principal objetivo a transferência de recursos, gera uma necessidade cada vez maior de evolução das tecnologias e processos de decisão voltados a concessão do crédito (CHAIA, 2003).

No mercado como um todo é imprescindível o uso de modelos que facilitem a concessão desse crédito. Por sua maior agilidade, igualdade e menor custo, esse tipo de análise tem sido empregada em larga escala no mercado, atualmente grande parte dos modelos aplicados tem como resultado uma probabilidade de inadimplência por parte do cliente atribuída através do *credit score*, geralmente levando em consideração um conjunto de variáveis preestabelecido (ARAÚJO, 2006) (MANOS, 2006).

Essa maneira tradicional de quantificar o risco de crédito dos clientes, pode ser justificada no processo de aquisição, porém quando falamos de análise comportamental, ou seja, posterior a entrada dos clientes, torna-se necessário não só entender a probabilidade e sim o tempo até ocorrência. Essa visão pode restringir o público alvo de alguma ação com o objetivo de conter a inadimplência em momentos de crise, facilitar a tomada de decisão e antecipar os movimentos de inadimplência, auxiliando na solvência das instituições financeiras e representando uma vantagem competitiva em relação aos demais *players* do mercado (CAO; VILAR; RIVERA, 2009).

4 REVISÃO LITERÁRIA

Um conceito inerente a transferência de recursos, entre um agente superavitário e um deficitário, é o de crédito. Para [Schrickel \(2000\)](#), o crédito é todo ato de vontade ou disposição de alguém destacar ou ceder, temporariamente, parte do seu patrimônio a um terceiro, com a expectativa de que esta parcela volte a sua posse integralmente.

O crédito pode ser atrelado a um processo natural da evolução do conceito de moeda, tendo os seus primeiros relatos na Babilônia e no Egito. A chegada da coroa portuguesa ao Brasil em 1808 e por consequência o crescimento da atividade econômica, principalmente nos portos, foram fatores determinantes para regulamentação e implementação de um Sistema Financeiro Nacional que é dividido através do ramo de atividade em três mercados: (i) Moeda, Crédito, Capitais e Câmbio (ii) Seguros, Previdência complementar e resseguro (iii) Previdência Fechada (Fundo de pensão) ([ABREU, 2017](#)) e ([CAMINHA, 2021](#)).

A rápida evolução do Sistema Financeiro Nacional como um todo, foi por muito tempo subestimada por uma estrutura de regulação branda.

(...)Em 1864, a ausência de mecanismos de controle/estabilização das emissões de títulos se desdobrou na primeira crise financeira do Brasil: a quebra da maior casa bancária do império, a José Alves Souto e Cia, levando à (também primeira) corrida bancária de depositantes brasileiros para resgatar seus ativos. ([CAMINHA, 2021](#))

Após algumas outras tentativas de regulamentação, a reforma bancária e o novo marco legal da Sistema Financeiro Nacional foram publicados em 1964, dispendo além dos fatores organizacionais, sobre a criação do Conselho Monetário Nacional (CMN) e do Banco Central do Brasil (Bacen), sendo esses os órgãos responsáveis respectivamente por normatizar e supervisionar o mercado de crédito no Brasil.

Esse mercado tem ganhado espaço e se especializado a medida que, a melhoria na alocação desses recursos, afeta diretamente o desenvolvimento, inovação e incentiva o crescimento da economia de consumo ([SHANG; SONG; WU, 2017](#)). No Brasil, após as diversas crises inflacionárias e a implementação do plano real, o mercado de crédito tem se tornado cada vez mais abrangente e competitivo através dos diversos produtos oferecidos.

Um dos produtos mais comercializados no país, impulsionado por um avanço de tecnologias de informação, urbanização acelerada e controle da hiperinflação ([COSTA; COSTA; OLIVEIRA, 2010](#)), é o cartão de crédito. Segundo pesquisa da Associação Brasileira de Empresas de Cartões de Crédito e Serviços [ABECS \(2022\)](#) a modalidade obteve crescimento de 42,2% no primeiro semestre de 2022, representando pela primeira vez o montante de R\$ 1 trilhão em pagamentos nesse período de tempo.

As instituições financeiras que oferecem o serviço de cartão de crédito tem um papel de intermediar a relação entre os poupadores e tomadores, de modo que o risco inerente dessa transferência de valor é assumido de maneira integral pela instituição financeira. O retorno relacionado a essa movimentação é descrito como *Spread* ou "prêmio pela perda esperada", na prática essa balança é controlada pela diferença entre o valor captado pelas instituições e o valor "vendido" no mercado de crédito (DANTAS, 2020).

Paralela a relação de certa "confiança" entre a instituição e o tomador, é necessário um controle de risco nos momentos de análise inicial e monitoramento futuro, com o intuito de prevenir a quebra natural do ciclo de crédito.

Diante as incertezas do mercado de crédito, uma questão é bastante evidente, a necessidade de exercer um controle sobre as possíveis perdas, estabelecendo-se assim uma ciência de medição de risco de crédito e, com ela, um mercado de análise de crédito (SEBBEN, 2021).

No Brasil, esse movimento de controle sobre o risco torna-se ainda mais necessário levando em consideração a "normalização" do ato de endividar-se e o senso comum de visualizar o crédito como um aumento instantâneo de renda (SOUZA, 2013).

Os modelos de crédito surgem para auxiliar essa demanda de controle do risco, Fernandes e Artes (2013), definem os modelos de pontuação de crédito como ferramentas importantes no processo de concessão de crédito, medindo o risco de crédito de um cliente potencial baseado de maneira ampla, em variáveis idiossincráticas e fatores macroeconômicos. Com o objetivo de segundo Chaia (2003), obter um conjunto de parâmetros que demonstrem a real chance do devedor honrar a dívida.

A execução desses modelos tem como resultado uma pontuação de crédito definida por Avery, Brevoort e Canner (2009) como uma tecnologia estatística que quantifica o risco de crédito apresentado por um mutuário em potencial ou atual. Auxiliando as organizações no ato de decidir sobre a concessão do crédito e posteriormente a monitorar os clientes em seu portfólio (THOMAS, 2000). A figura 1 apresenta o diagrama de decisão para concessão de crédito.

Figura 1 – Fluxo da concessão de crédito utilizando um modelo de *Credit Scoring*.



Fonte: (CHAIA, 2003), p.30

A tomada de decisão utilizando como parâmetro a pontuação de crédito decorrente dos resultados obtidos na implementação dos modelos de crédito é um fator dependente da evolução dos computadores e da tecnologia em geral, dessa forma, apesar do seu crescimento astronômico, ambos são recentes na história (ANDERSON, 2007).

A convocação para a segunda guerra mundial minou a mão de obra especializada em conceder crédito e exaltou a necessidade de modernizar o setor. Os primeiros processos de decisão baseados em uma pontuação de crédito foram desenvolvidos por David Durand em 1941 com base em uma pesquisa feita na área da biologia pelo cientista britânico Ronald Aylmer Fisher, expandindo-se apenas no pós guerra, mesmo com a resistência de algumas empresas na automatização da tomada de decisão por parte dos computadores e a enorme dificuldade da aplicação dos cálculos estatísticos nos softwares disponíveis na época (ANDERSON, 2007).

A chegada do produto cartão de crédito em meados de 1960, o acordo de Basileia II assinado em 2004 e a crise de 2008, foram eventos chave para implementação em escala geral e evolução dos modelos de crédito com o principal objetivo de antecipar eventos de perdas e inadimplência dentro das instituições financeiras.

(...)Recently, several financial crises, such as credit and cash card crises in Taiwan and subprime mortgage crisis in US, have not only caused a severe recession in domestic economy but resulted in the global financial disaster. (HSIEH; LEE; LEE, 2011)

De maneira concomitante, a oficialização do IFRS 9 foi responsável por ampliar o uso desses modelos de crédito e difundir a visão entre as pequenas e médias empresas que passaram a ser estimuladas a desenvolver modelos de mensuração e classificação de risco. Além dos requisitos propostos em acordos globais e leis específicas de cada Sistema Financeiro, a demanda do mercado por mais igualdade, velocidade e redução de custos nos diversos processos que envolvem a concessão do crédito e retenção de clientes, transforma os modelos de crédito em parte essencial para a sobrevivência da instituição financeira (DEMIRGUC-KUNT; KLAPPER; SINGER, 2017).

É importante frisar que além da concessão de crédito assertiva a empresa tem o papel de manter o portfólio alinhado com as expectativas de perda e inadimplência que podem variar de acordo com o produto ofertado e a área de atuação da instituição. Como os modelos de crédito aplicados na originação levam em consideração os dados demográficos e de histórico de crédito no mercado geral, essa originação pode variar de acordo com as mudanças realizadas na política de crédito de cada instituição, ou seja, podemos ter diversas Safras com características e personalidades diferentes. Nesse sentido, os modelos conhecidos como *behaviour scoring* têm o objetivo de englobar o comportamento do cliente dentro da instituição e os fatores mutáveis citados anteriormente, gerando como resultado um fator conhecido pelo mercado como escore comportamental (SEBBEN, 2021).

Segundo Hsieh, Lee e Lee (2011), esses modelos que visam quantificar o comportamento do cliente ganharam cada vez mais atenção devido à explosão da concorrência nos últimos anos, permitindo às empresas tomarem decisões mais rápidas e assertivas.

Um dos produtos mais abrangentes, no sentido de quantidade de oferta elevada no mercado, em grande parte das instituições financeiras é o cartão de crédito, nele o *behavior scoring* é utilizado, entre outros processos, para auxiliar na antecipação do movimento de inadimplência, na tomada de decisão relacionada ao aumento de limite (majoração) e na disponibilidade de um percentual a mais de crédito fora do valor pré-aprovado na concessão (overlimit), sendo esses dois últimos critérios parte das conhecidas políticas de manutenção (LI; ZHONG, 2012).

É comum que as empresas busquem por melhorias nos seus modelos de crédito gerais, visando a redução de custos provenientes da inadimplência através da escolha correta e acompanhamento efetivo dos clientes, nesse sentido o ramo da Análise de sobrevivência vem ganhando espaço no mercado de crédito, na medida em que torna possível modelar o tempo até a inadimplência e não apenas a probabilidade de ocorrência do evento, conhecida na literatura como Probabilidade *Default* ou PD. A literatura sobre análise de sobrevivência sempre esteve historicamente ligada a áreas como a medicina, psicologia e engenharia de produção (COLOSIMO; GIOLO, 2006), com sua aplicação dentro do mercado de crédito iniciado por Narain (1992) e desenvolvida por Stepanova e Thomas (2002), que comparou os modelos Weibull, Exponencial e de Cox com a técnica mais usada na época, a Regressão logística, tendo como descoberta que os modelos de Análise de sobrevivência, sejam eles paramétricos ou semi paramétricos, são na maioria das vezes superiores a regressão logística.

5 METODOLOGIA

5.1 Material

5.1.1 Base de dados

A Base de dados utilizada nesse estudo é composta por clientes de uma empresa multinacional varejista, que atua em quase todas as regiões do Brasil, oferecendo o produto cartão de crédito em suas lojas. De maneira a preservar a imagem da empresa citada, o estudo se dirige a tal como empresa A. O extenso volume de dados da base, tornou necessário a escolha aleatória de 100.000 observações, além da retirada de algumas variáveis que feriam a Lei Geral de Proteção de dados(LGPD).

Inicialmente foram selecionados os clientes com entrada a partir do ano de 2000, levando em consideração o reduzido volume de clientes com data de entrada anterior. Sendo observado o horizonte temporal de 22 anos, onde 10% entrou em inadimplência dentro desse período.

5.2 Recursos

Para a análise descritiva dos dados e o ajuste do modelo paramétrico, foi utilizado o software R, versão 4.1.2 (R CORE TEAM, 2017), tendo como principais pacotes utilizados: survminer, survival, gmodels, dplyr, tidyverse, ltm e nortest.

5.3 Métodos

5.3.1 Seleção de Variáveis

Os modelos comportamentais aplicados ao cartão de crédito são construídos com base em informações de crédito internas (utilização do cartão, saldo devedor, histórico de inadimplência, comprometimento do cartão, ...), informações de perfil (idade, renda e limite da concessão), informações de bureau que servem de forma complementar ao resultado interno, além do escore obtido na própria originação, compõe a maioria das variáveis.

A mineração desses dados torna-se um fator importante, tendo como objetivo a escolha correta das variáveis que serão utilizadas no modelo. De maneira geral essa decisão depende do objetivo de cada instituição e não existe uma regra que norteie esse processo. Assim como em outros modelos aplicados no mercado, podemos dividir o processo de seleção de variáveis em dois grandes grupos: O de maneira subjetiva, onde a instituição define as variáveis que serão selecionadas para o modelo e a objetiva, cujo a seleção é definida através de métodos de separação de variáveis ou redução de dimensão do banco de dados (MAFIO, 2007).

No presente estudo foi aplicado o método de seleção subjetivo, utilizando dados de perfil e o escore de crédito obtido com o modelo de crédito da aquisição.

5.3.2 Análise de sobrevivência

O conjunto de técnicas aplicadas que formam a análise de sobrevivência tem o objetivo de, segundo [Borges \(2014\)](#), analisar os tempos de vida dos indivíduos desde o seu momento de entrada no estudo, até o momento em que ocorre o evento de interesse, que deve ser definido na partida.

Um conceito derivado da escolha do evento de interesse dentro da análise de sobrevivência é o de tempo de falha, que consiste na subtração entre a data de início do estudo e o momento em que o evento ocorre, sendo em sua grande maioria medidos através de tempo real/relógio. O estudo dessa variável aleatória positiva (T) é o ponto principal da análise de sobrevivência ([COLOSIMO; GIOLO, 2006](#)). Nos modelos comportamentais aplicados ao mercado de crédito, o tempo de falha em vias gerais é entendido como o tempo até o cliente entrar em processo de inadimplência.

As duas funções chave para estudos que envolvem a análise de sobrevivência, são função de sobrevivência e a função de risco. A função de sobrevivência define a probabilidade de uma observação com tempo de falha T ultrapasse um determinado tempo t e não experimente a falha. Conhecida como probabilidade de sobrevivência, essa função tem como características

$$S(0) = 1$$

e

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

sendo expressa por

$$S(t) = P(T > t)$$

A função de risco, mede o risco da falha ocorrer em um determinado intervalo de tempo $[t_1, t_2)$, tendo como resultado uma taxa. Ao assumir um valor pequeno para o intervalo entre os tempos de interesse, definimos a taxa de falha instantânea em um determinado tempo. A característica da função de risco é não possuir nível superior definido, ao contrário da função de sobrevivência ([MOORE, 2016](#)).

A função de risco levando em consideração o tempo de falha T pode ser expressa por

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

em que, Δt representa a diferença entre os tempos determinados e t o tempo inicial. Sendo sua acumulada dada por

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

Uma técnica muito utilizada na análise de sobrevivência é observar a taxa de falha instantânea em relação ao tempo, via de regra existem três resultados esperados: (i) função crescente, (ii) função constante e (iii) função decrescente. No presente estudo, cuja falha é definida como a inadimplência do cliente, é comum encontrar uma função crescente, ou seja, a medida que o tempo aumenta, o cliente tem um risco maior de não pagar o cartão de crédito, devido a permanência de exposição ao risco e aos fatores macroeconômicos/pessoais serem mutáveis.

A propriedade amplamente utilizada nos estudos que envolvem a análise de sobrevivência é a de relação direta entre as funções apresentadas, como exemplo, através da função $S(t)$ é possível encontrar $\lambda(t)$ e $\Lambda(t)$.

Ao se trabalhar com a variável tempo é imprescindível determinar de forma categórica o período observado para o estudo, a fim de viabilizar o processo de análise. Com essa restrição temporal surgem as censuras, que representam as observações incompletas ou inexistentes sobre o evento de interesse. [Colosimo e Giolo \(2006\)](#) destaca dois fatores que justificam a utilização integral de dados censurados na Análise de sobrevivência:

- Mesmo sendo incompletas, as observações censuradas nos fornecem informação do tempo de vida;
- A omissão das censuras no cálculo das estatísticas de interesse pode acarretar em conclusões viciadas.

As censuras podem ser separadas em três categorias, (i) censura à direita, caracterizado por ter o tempo de ocorrência do evento, mesmo que de forma parcial, computado depois do início do estudo, (ii) censura à esquerda, ocorre quando o tempo de vida é inferior ao início do estudo, (iii) censura intervalar, gerada a partir da falta de certeza relacionada ao tempo em que o evento de interesse ocorreu ([COLOSIMO; GIOLO, 2006](#)).

A maioria das aplicações de Análise de sobrevivência no desenvolvimento dos modelos de crédito incorre na censura a direita, devido a ampla gama de possibilidades para o rompimento da relação entre a instituição e o cliente, como exemplo: cancelamento involuntário/voluntário,

falecimento, roubo, tentativa de fraude, inatividade, transferência de titularidade e vencimento do cartão. Dessa forma o número de clientes que entra em inadimplência é reduzido, ocasionando grande volume de censura a direita que pode ser apresentada em três subdivisões:

- Tipo 1: O estudo é finalizado após um tempo preestabelecido;
- Tipo 2: O estudo é finalizado após ter ocorrido um volume n de eventos de interesse;
- Aleatória: O indivíduo é retirado do estudo sem ter ocorrido o evento de interesse.

Cao, Vilar e Rivera (2009), descreve a peculiaridade das censuras à direita na aplicação das técnicas de Análise de sobrevivência nos modelos de crédito da seguinte forma:

(...)Na prática, como a proporção de créditos inadimplentes é pequena, a proporção de dados censurados é grande, o que pode causar baixo desempenho dos métodos estatísticos. Por outro lado, o tamanho da amostra é tipicamente muito grande. Isso alivia de alguma forma o problema da grande proporção de censura.

Em um estudo de análise de sobrevivência os conceitos de tempo de falha(t_i) e censura(α_i) são apresentados em forma de coordenadas bipolares dadas por (t_i, α_i) .

5.3.3 Estimador de Kaplan-Meier

Um dos primeiros passos em um estudo que utilize dados de sobrevivência, é entender os tempos de sobrevivência das observações, dessa forma o estimador de kaplan-meier é responsável por descrever a função de sobrevivência ao longo do tempo, aplicando os conceitos de independência de eventos e probabilidade condicional a fim englobar as informações censuradas, que não conseguem ser ajustadas a uma análise descritiva gráfica padrão (COLOSIMO; GIOLO, 2006) e (CARVALHO et al., 2011).

Primeiro devemos considerar que a função $S(t)$ pode ser escrita através de probabilidade condicional, levando em consideração que t_j é o tempo em que ocorreram as $k(\leq n)$ falhas

$$S(T_j) = (1 - q_1)(1 - q_2)\dots(1 - q_j)$$

E q_j , é dado por

$$q_j = P(T \in [t_{j-1}, t_j) | T \geq t_{j-1})$$

representando a probabilidade de uma observação falhar no intervalo de tempo $[t_{j-1}, t_j)$ dado que a observação sobreviveu até t_{j-1} . Ao escrever a função de sobrevivência de maneira condicional o estimador de Kaplan-Meier passa a ter a função de estimar o q_j , através da equação

$$\hat{q}_j = \frac{\text{N}^\circ \text{ de falhas em } t_j}{\text{N}^\circ \text{ de observações sob risco em } t_{j-1}}, \text{ para } j \leq 1$$

que parte da adaptação da função utilizada para estimar a função de sobrevivência de dados não censurados, dada por

$$\hat{S}(t) = \frac{\text{N}^\circ \text{ de observações que não falharam até o tempo } t}{\text{N}^\circ \text{ total de observações no estudo}}$$

A expressão geral do estimador de Kaplan-Meier pode então ser definida como

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right)$$

Onde,

- t_j é o tempo de falha;
- d_j é o número de falhas;
- n_j é o número de indivíduos que estão expostos ao risco no tempo t_j .

sendo essa uma função escada em que seus degraus descrevem os tempos de falha observados. Esse estudo partiu da construção do estimador de Kaplan-Meier com o objetivo de entender a relação entre as variáveis selecionadas e o tempo até a inadimplência dos clientes, em alguns momentos utilizando o teste de *log-rank* para evidenciar a diferença entre as curvas.

5.3.4 Modelos Paramétricos

O modelo de *Kaplan-Meier* apresenta de maneira descritiva os dados de sobrevivência e é utilizado como análise preliminar dos dados. Como o objetivo do presente estudo é entender a influência de cada variável selecionada no processo de inadimplência, foi utilizado o modelo paramétrico ou probabilístico, que tem a finalidade de vincular o tempo de sobrevivência da observação (T) a uma variável explicativa (covariável), através de uma distribuição de probabilidade conhecida (CARVALHO et al., 2011).

Existem diversas distribuições amplamente utilizadas em análises estatísticas. Quando tratamos de uma variável aleatória associada ao tempo, algumas distribuições mostraram-se mais adequadas por serem positivas e com histórico de aplicação na área de análise de dados de sobrevivência, entre elas: Exponencial, Weibull e Log-normal (COLOSIMO; GIOLO, 2006).

5.3.4.1 Distribuição Exponencial

O modelo exponencial é considerado um dos mais simples para descrever o tempo até a falha de um indivíduo, possui um único parâmetro e tem a característica da sua função de risco (taxa de falha) ser constantes (MOORE, 2016) e (CARVALHO et al., 2011). Sua função densidade de probabilidade é dada por

$$f(t) = \frac{1}{\alpha} \exp - \left(\frac{t}{\alpha} \right), \quad t \geq 0$$

O parâmetro $\alpha \geq 0$, determinado na função, tem a mesma unidade do tempo de sobrevivência t e representa o tempo médio de sobrevivência das observações.

A função de sobrevivência e a função de risco da distribuição exponencial são dadas por:

$$S(t) = e^{-(at)}$$

$$\lambda(t) = \frac{1}{\alpha} \quad \text{para } t \geq 0 \quad (5.1)$$

A propriedade de "falta de memória" da distribuição exponencial fica evidente na função constante de risco (5.1) (CARVALHO et al., 2011).

5.3.4.2 Distribuição Weibull

A Weibull é uma das distribuições mais utilizadas ultimamente. Devido a sua função de risco assumir característica crescente, decrescente ou contante, sua aplicação é aderente a diversos cenários que envolvem estudo de tempo de sobrevivência (MOORE, 2016). A função densidade de probabilidade atribuída a uma variável aleatória T , que representa o tempo de sobrevivência dos indivíduos, com distribuição de probabilidade Weibull é descrita como

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp - \left(\frac{t}{\alpha} \right)^\gamma, \quad \text{para } t \geq 0$$

Em que ambos os parâmetros da distribuição (α, γ) são positivos. A função de sobrevivência e a função de risco da distribuição Weibull são dadas respectivamente por

$$S(t) = \exp - \left(\frac{t}{\alpha} \right)^\gamma$$

$$\lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}$$

Quando $\gamma = 1$ tem-se

$$\lambda(t) = \frac{1}{\alpha}$$

Dessa forma, pode-se assumir a distribuição exponencial como um caso particular da distribuição Weibull.

5.3.4.3 Distribuição Log-Normal

Quando a variável aleatória contínua T tem seu logaritmo normalmente distribuído ela assume a característica de uma Log-normal, ou seja, uma variável aleatória com distribuição Log-normal pode ser analisada segundo uma distribuição normal, desde de que, considere-se o logaritmo dos dados apresentados (COLOSIMO; GIOLO, 2006) e (CARVALHO et al., 2011).

Na função de densidade da distribuição Log-normal representada na equação 5.2, o parâmetro μ é entendido como a média do logaritmo do tempo de sobrevivência e σ o seu desvio-padrão.

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2, \quad \text{para } t \geq 0 \quad (5.2)$$

As funções de sobrevivência e de risco para uma distribuição Log-normal são representados respectivamente por:

$$S(t) = \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)$$

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

Em que Φ é a função de distribuição acumulada da normal padrão.

5.3.4.4 Escolha do modelo e estimação dos parâmetros

O fator chave para análise e adequação do modelo paramétrico aos dados de sobrevivência é a escolha correta do modelo de regressão ao qual possibilita observar o efeito das covariáveis no tempo de sobrevivência. No presente estudo foram utilizados métodos gráficos e o teste da razão de verossimilhança, para definir o modelo a ser aplicado (COLOSIMO; GIOLO, 2006).

O primeiro método gráfico consiste na comparação da curva do estimador de Kaplan-Meier e o modelo proposto, levando em consideração a estimação dos seus parâmetros e por consequência sua função de sobrevivência. Dessa forma, quanto mais próximo as curvas de sobrevivência estiverem da estimativa de Kaplan-Meier, melhor o modelo.

O segundo método gráfico busca linearizar a função de sobrevivência utilizando as distribuições descritas e comparar com uma reta, esse método geralmente determina de maneira

mais evidente seu resultado a medida que as violações a reta de comparação ficam mais evidente e elimina parte da análise "subjativa" que pode existir no método 1.

Como nem sempre é possível tomar alguma decisão com base nas análises gráficas descritas, foi plausível introduzir no estudo um componente que utiliza-se do teste de hipótese, retirando o fator subjetivo da análise sujeita a interpretações (SAUDERS, 2000) e (COLOSIMO; GIOLO, 2006).

A hipótese nula H_0 nesse sentido afirma que, o modelo de interesse é adequado. Partindo de um modelo generalizado, geralmente utilizando a distribuição gamma que tem como caso especial os demais modelos apresentados, define-se o valor do logaritmo da sua função de verossimilhança e posteriormente aplica-se esse processo para o modelo de interesse, isto é

$$TRV = -2 \log \left[\frac{L(\hat{\theta}_M)}{L(\hat{\theta}_G)} \right] = 2[\log L(\hat{\theta}_M) - \log L(\hat{\theta}_G)]$$

Após a escolha da distribuição que mais representa os dados de sobrevivência é necessário estimar os parâmetros dessa distribuição com base nos dados amostrais. Na estatística existem diversos métodos para estimação de parâmetro, porém o de máxima verossimilhança é o que melhor consegue incorporar os dados censurados (COLOSIMO; GIOLO, 2006).

O método consiste na escolha dos melhores parâmetros para a amostra específica, ou seja, determina quais os valores do parâmetros que me retornam a maior probabilidade de obter como resultado a amostra em questão. Incorporando através de funções específicas cada tipo de censura e as observações não-censurados.

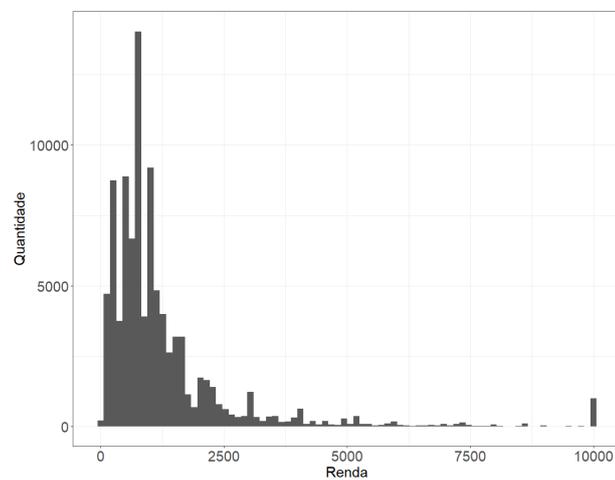
Além da escolha do modelo, torna-se necessário o estudo dos resíduos, com o objetivo de entender o quanto o modelo foi preciso nos resultados. A maneira mais comum é analisar o ajuste final do modelo com base nos resíduos de Cox-Snell, que segundo Colosimo e Giolo (2006), são obtidos através da estimativa dos erros que vem de uma população homogênea, devendo seguir uma distribuição exponencial padrão se o modelo for adequado.

6 RESULTADOS

6.1 Análise Descritiva e Exploratória

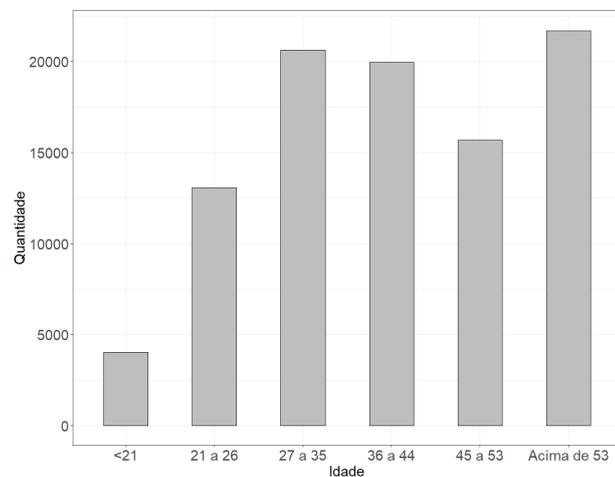
O perfil de renda e idade na aquisição dos clientes segue um padrão bem estabelecido, as figuras 2 e 3 mostram uma inclinação a originação de clientes com renda até R\$ 2.500 e idade acima de 21 anos, com média de 41 anos e desvio padrão de 14 anos. Os fatores que podem influenciar esse resultado são a política de crédito e o seguimento de varejo cujo a empresa A está inserida.

Figura 2 – Perfil de renda dos clientes da empresa A.



Fonte: Elaboração própria

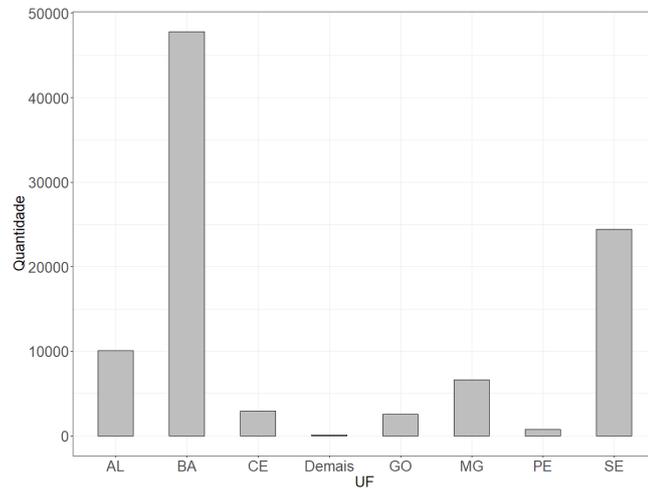
Figura 3 – Perfil de idade na aquisição dos clientes da empresa A de 2000 a 2022



Fonte: Elaboração própria

Grande parte dos clientes que tem o produto cartão de crédito estão contidos na região nordeste, a região centro-oeste está representada pelos estados de Minas Gerais e Goiás (figura 4).

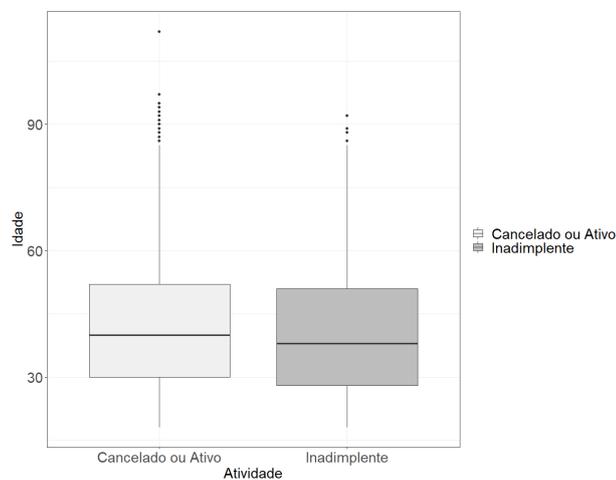
Figura 4 – Informação de UF, obtido na aquisição dos clientes da empresa A de 2000 a 2022



Fonte: Elaboração própria

Para fins de comparação os clientes cancelados e ativos foram agrupados, a marcação de cancelados leva em consideração fatores como: falecimento, cancelado voluntário/involuntário e troca de titularidade. A figura 5 mostra certa paridade na média e concentração das idades entre os clientes que entraram em inadimplência e os que estão cancelados/ativos. O volume de valores atípicos para os clientes classificados em cancelados/ativos tende a ser maior devido a abrangência dos falecidos.

Figura 5 – Comparação entre perfil de idade na aquisição dos clientes cancelados/ativos e inadimplentes.



Fonte: Elaboração própria

A classificação de renda retrata o resultado do perfil de idade dos clientes, a tabela 1 evidência uma maior disponibilidade de renda para clientes com idade superior a 36 anos, chegando a representar 87,2% dos clientes com renda acima de R\$ 3.600.

Tabela 1 – Frequência relativa entre a idade e renda dos clientes da empresa A de 2000 a 2022.

Idade	Renda			
	Até um S.M.	R\$ 1.200 A 2.400	R\$ 2.400 A 3.600	Acima de R\$ 3.600
<21	6,0%	0,6%	0,5%	0,5%
21 a 26	18,4%	4,4%	3,2%	1,7%
27 a 35	24,6%	16,8%	15,2%	10,5%
36 a 44	20,4%	22,7%	22,7%	20,1%
<45 a 53	14,3%	20,3%	22,9%	23,2%
>53	16,3%	35,2%	35,5%	44,0%
Total	100,0%	100,0%	100,0%	100,0%

Fonte: Elaboração própria

A empresa A em questão oferta seu cartão com base em um modelo de crédito que utiliza a regressão logística para obter um escore de crédito (*Credit Scoring*), que é utilizado para categorizar os clientes de Baixíssimo até Alto risco, a partir da tabela 2.

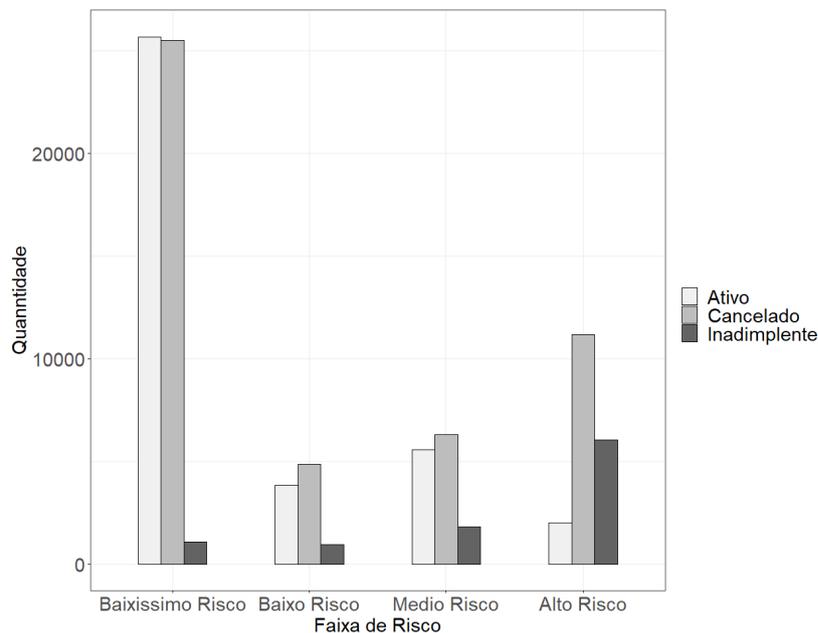
Tabela 2 – Definição da faixa de risco dos clientes com base no modelo de crédito aplicado na empresa A.

Faixa de risco	<i>Credit Scoring</i>
Baixíssimo	851 a 1.000
Baixo	501 a 850
Médio	251 a 500
Alto	0 a 250

Fonte: Elaboração própria

A faixa de risco dos clientes discrimina bem o resultado de inadimplência, como evidenciado na figura 6. Em sua grande maioria os clientes inadimplentes tem a faixa de risco da aquisição atrelada ao Alto risco.

Figura 6 – Atividade dos clientes em relação a Faixa de risco da obtida na aquisição dos clientes entrantes de 2022.

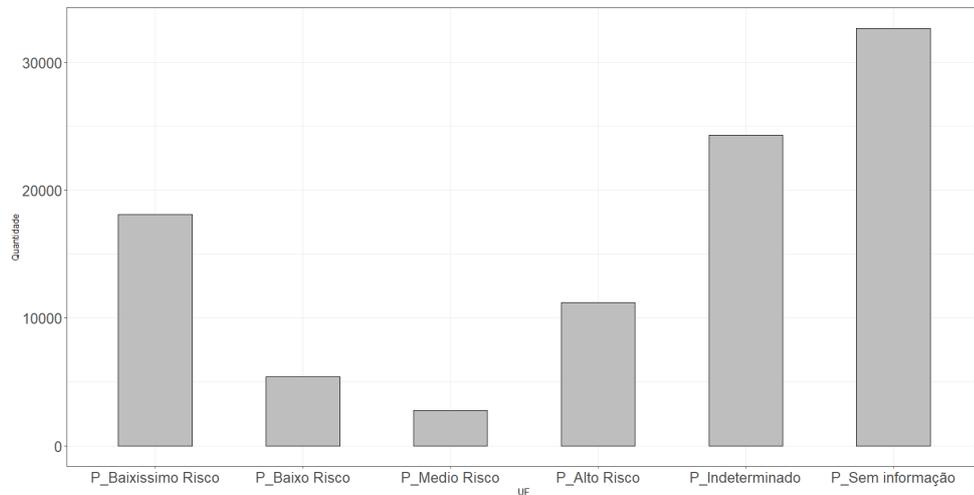


Fonte: Elaboração própria

O modelo comportamental padrão aplicado na empresa A, utiliza um modelo de regressão que retorna um escore, sendo posteriormente classificado em: Baixíssimo Risco, Baixo Risco, Médio Risco, Alto Risco, Sem informação e Indeterminado. Nesse estudo para diferenciar a classificação dos modelos de *behavior scoring* e *credit scoring* aplicados na empresa, vamos utilizar o prefixo "P" para determinar o modelo comportamental "padrão".

A figura 7 apresenta a quantidade de clientes por classificação de risco, as observações classificadas em "Sem informação" representam os clientes cancelados acima de seis meses, já as observações classificadas com "Indeterminado" são referente a clientes cancelados a menos de seis meses e ativos que estão a menos de seis meses no portfólio ou não tem histórico de compra definido. A maioria dos clientes possui classificação de risco ,no modelo comportamental, como indeterminado ou sem informação, o que está relacionado a parte dos clientes ter cancelamento ou não ter informação de seis meses seguidos de compra dentro da empresa A.

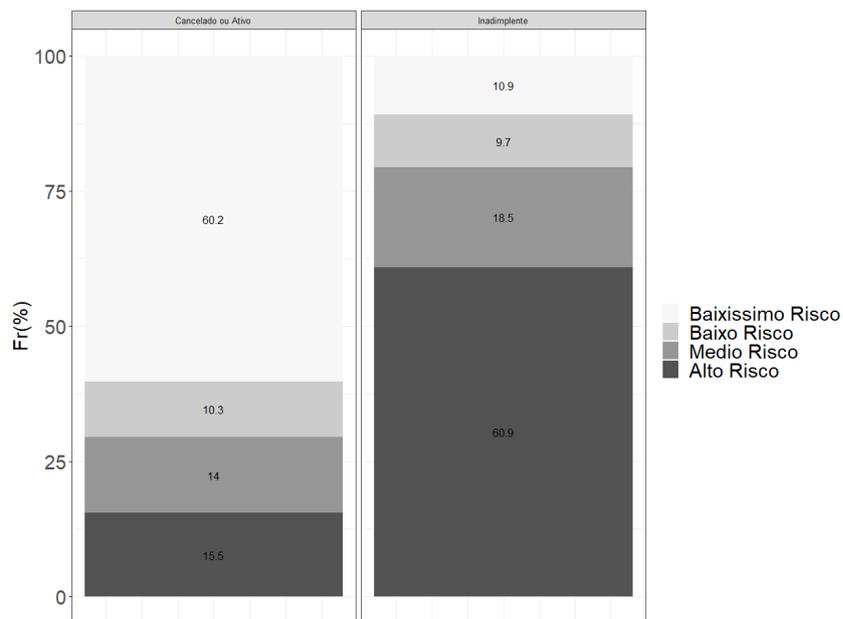
Figura 7 – Classificação de risco aplicada no modelo comportamental existente da empresa.



Fonte: Elaboração própria

A interação entre o resultado do modelo de crédito aplicado na aquisição e a atividade do cliente dentro da empresa A é descrito na figura 8, os clientes de alto risco representam 60,9% dos inadimplentes, enquanto a maioria dos cancelados ou ativos está na faixa de Baixíssimo risco.

Figura 8 – Perfil de idade dos clientes da empresa A



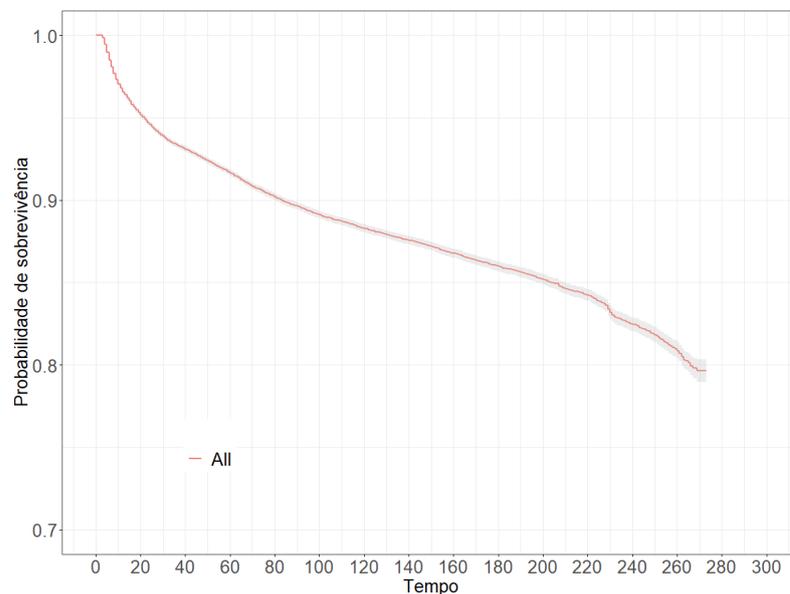
Fonte: Elaboração própria

6.2 Estimador de Kaplan-Meier

Além da estatística descritiva, utilizou-se o estimador de Kaplan-Meier, para traçar a relação entre o evento de interesse (inadimplência) e o tempo, levando em consideração as covariáveis selecionadas : Idade, Renda, Classificação Risco e Unidade Federativa (UF).

A figura 9 representa a curva de sobrevivência estimada através de Kaplan-Meier para todos os clientes. É possível notar uma redução para probabilidade de sobrevivência nas primeiras 25 semanas e posteriormente a partir das 230 semanas.

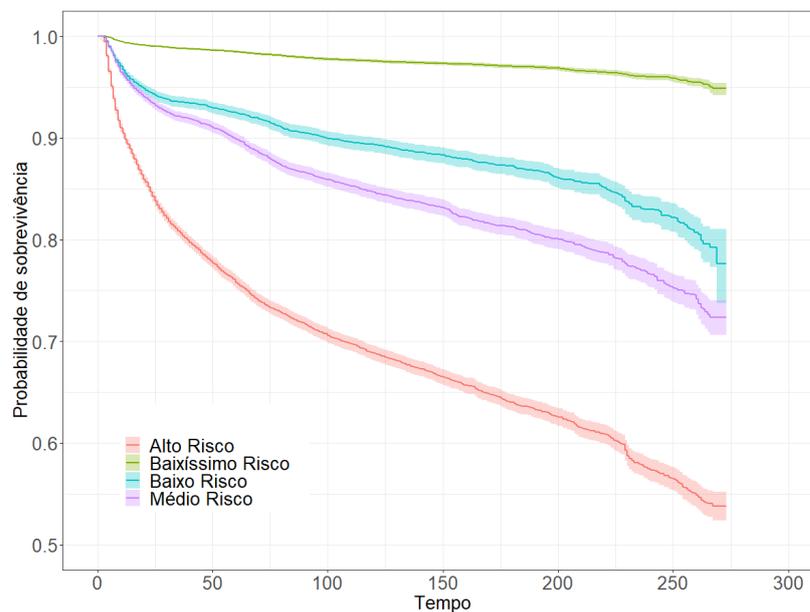
Figura 9 – Estimativa de Kaplan-Meier para a sobrevivência, em semanas, dos clientes que entraram em inadimplência.



Fonte: Elaboração própria

A figura 10 apresenta a curva de sobrevivência por classificação de risco. A divergência entre o comportamento de inadimplência fica evidente nas primeiras semanas, com o público Alto risco se distanciando em relação aos demais. O comportamento do público Alto Risco pode explicar a queda até a 25ª semana identificada no gráfico 9, a medida que a inadimplência ocorre de forma mais contundente nesse período, diferente das outras classificações.

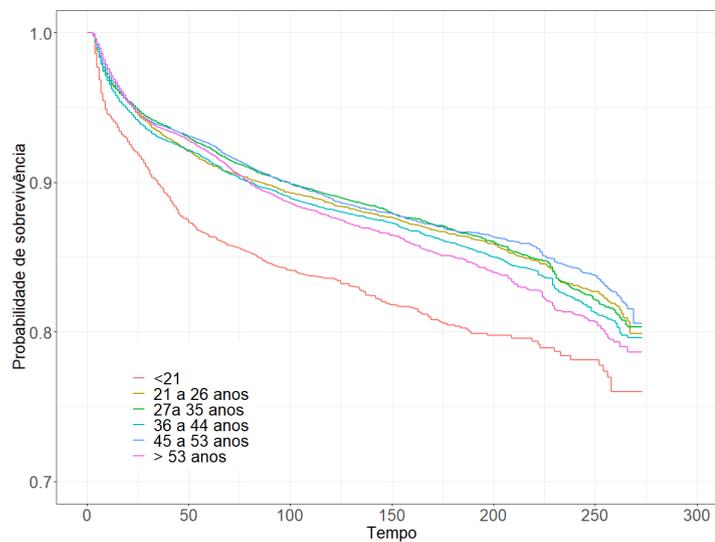
Figura 10 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Faixa de risco na aquisição.



Fonte: Elaboração própria

A idade dos clientes foi agrupada a fim de traçar as curvas de sobrevivência, a figura 11 apresenta um cenário de certa igualdade entre as idades acima de 21 anos e os clientes com idade abaixo de 21 anos se distanciando desde as primeiras semanas. O teste de log-rank foi aplicado estratificando os dados em clientes com idade superior ou igual a 21 anos e inferior a 21 anos, tendo como resultado a estatística de teste qui-quadrado de 100 com 1 grau de liberdade e o p-valor de $2e^{-16}$, ou seja, temos evidências suficientes para dizer que há uma diferença estatística significativa na curva de sobrevivência dos dois grupos. Dessa forma, os clientes com idade acima de 21 anos não discriminam bem o tempo até inadimplência e devem ser observadas outras características de maneira conjunta, para viabilizar a tomada de decisão na empresa A.

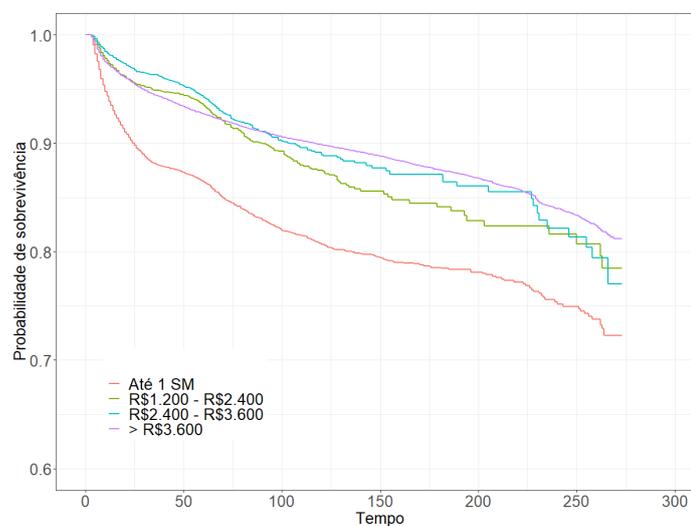
Figura 11 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Idade.



Fonte: Elaboração própria

A renda dos clientes está descrita através do estimador de Kaplan-Meier na figura 12, o público com renda até um salário mínimo tem caimento na curva de sobrevivida acelerado em relação aos demais. No tempo $t = 50$ semanas, a sobrevivida do grupo já atinja 87,5%, a titulo de comparação o grupo com renda entre R\$ 1.200 e R\$ 2.400 atingiu esse resultado apenas no tempo $t = 125$ semanas. Assim como no estimador aplicado a idade dos clientes, o grupo com renda acima de um salário mínimo não apresenta grande distinção no tempo de sobrevivida.

Figura 12 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Renda.

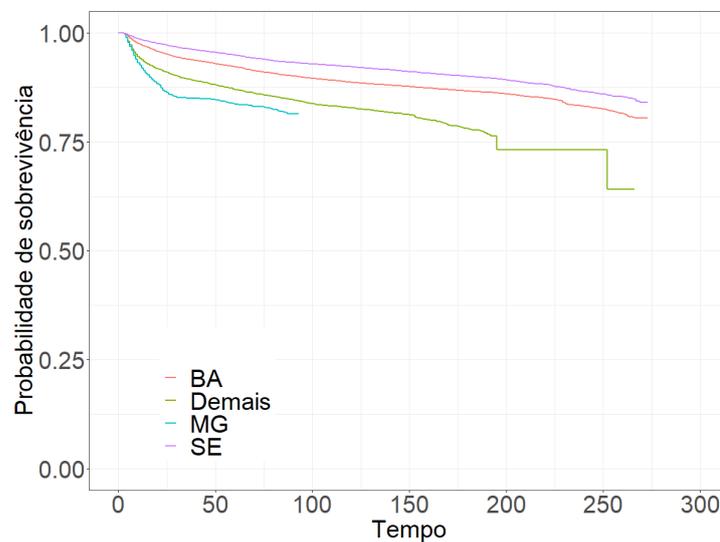


Fonte: Elaboração própria

Para fins de simplificar a visualização dos dados e montagem do modelo, algumas unidades federativas foram agrupadas devido ao reduzido volume de observações, a figura 13 mostra um comportamento parecido entre as curvas de sobrevivência dos estados de Bahia e Sergipe, além de um comportamento primário diferente para Minas Gerais.

O tempo em que foi inaugurada a empresa A em Minas Gerais pode ser um fator determinante para esse caimento nas primeiras semanas, levando em consideração a mudança recente do cenário de inadimplência no país esse impacto pode ser maior em "Safras" que estão em construção.

Figura 13 – Estimativa de Kaplan-Meier para a sobrevivência dos clientes que entraram em inadimplência, com base na covariável Unidade Federativa.



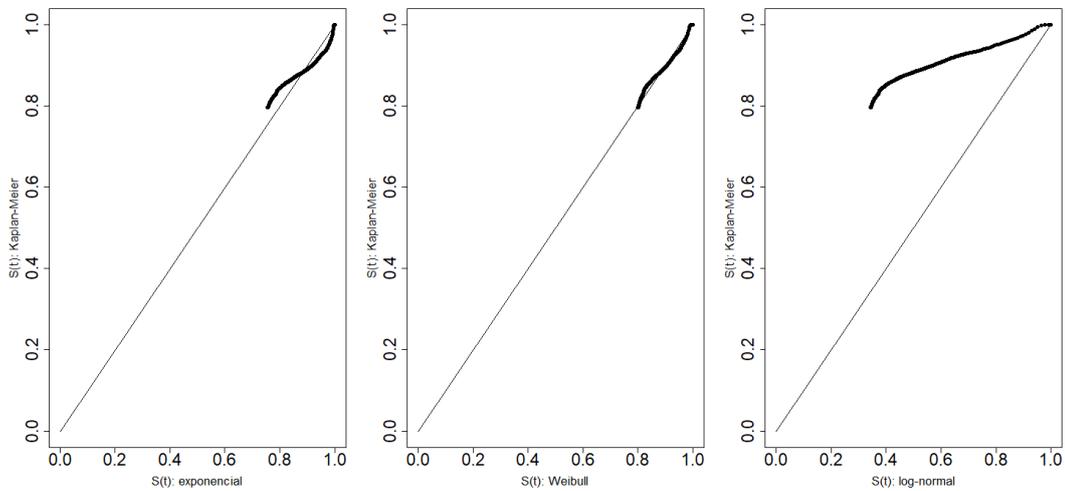
Fonte: Elaboração própria

6.3 Ajuste do modelo probabilístico

O primeiro estágio para obtenção do modelo probabilístico é entender a distribuição de probabilidade que devemos aplicar, para isso, serão utilizadas as técnicas apresentadas na seção 5, sendo testada as distribuições Exponencial, Weibull e Log-normal por terem um repertório nos estudo de análise de sobrevivência.

O primeiro método utilizado compara, a função de sobrevivência obtida através dos modelos gerais das distribuições com o estimador de Kaplan-Meier, levando em consideração o evento de inadimplência. A partir da figura 14 é possível observar uma tendência maior para utilização do modelo Weibull, seguindo bem alinhado a reta, por outro lado o modelo log-normal apresentou o resultado menos satisfatório.

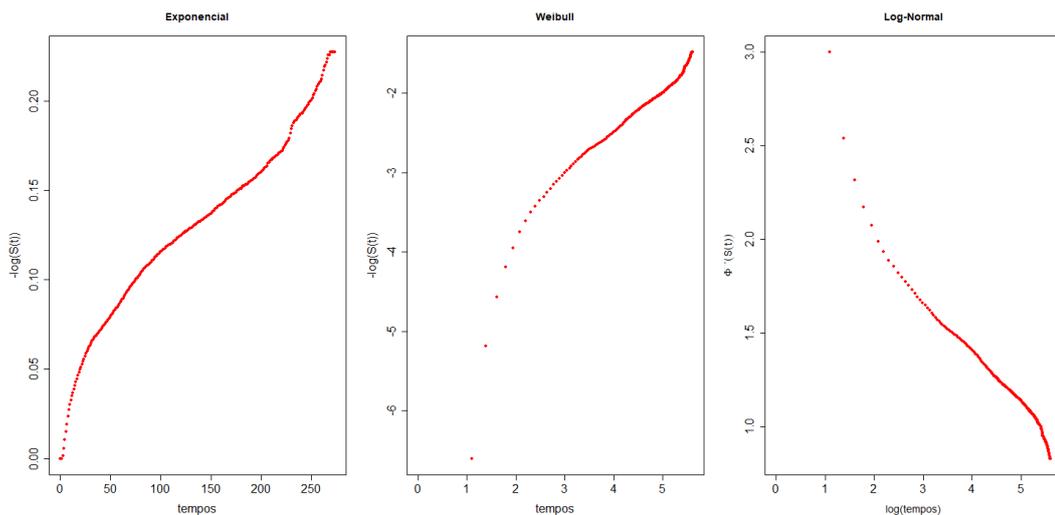
Figura 14 – Gráficos das Curvas de sobrevivência estimadas por Kaplan-Meier *versus* as sobrevivências estimadas pelos modelos Exponencial, de Weibull e log-normal.



Fonte: Elaboração própria

O segundo método consiste em linearizar a função de sobrevivência, A figura 15 não evidencia nenhuma grande discrepância da disposição dos dados em relação a uma reta, inclusive com o modelo de Weibull sendo o pior ajustado em relação aos demais, resultado divergente do apresentado no método 1.

Figura 15 – Gráficos t *versus* $-\log(\hat{S}(t))$, $\log(t)$ *versus* $\log(-\log(\hat{S}(t)))$ $\Phi^{-1} \log(t)$.

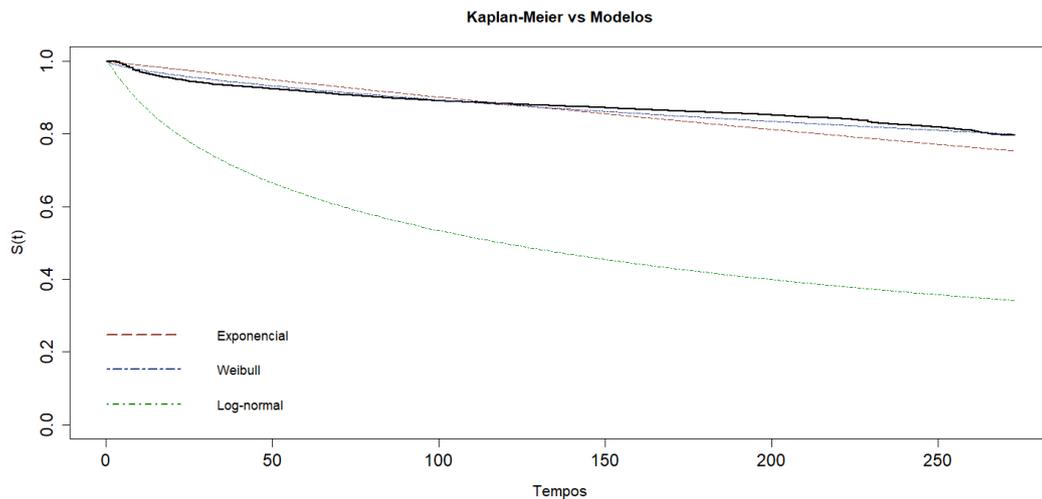


Fonte: Elaboração própria

O ultimo método gráfico utilizado para definir a escolha do modelo de probabilidade consiste em comparar a curva de sobrevivência do estimador de Kaplan-Meier e as curvas traçadas com os modelos probabilísticos. Assim como no método 1, os modelos Exponencial e

Weibull estão alinhados a curva de sobrevivência que foi estimada com Kaplan-Meier e o modelo log-normal segue um conjuntura diferente das demais curvas.

Figura 16 – Curva de sobrevivência estimadas por Kaplan-Meier *versus* as curvas de sobrevivência estimadas pelos modelos Exponencial, de Weibull e log-normal em relação ao tempo, com base no evento de inadimplência .



Fonte: Elaboração própria

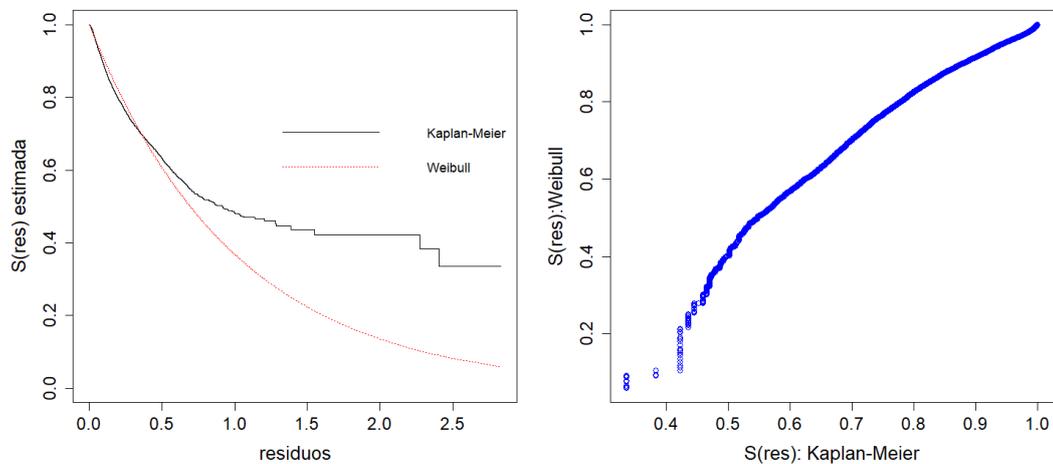
Como nem sempre a análise gráfica retorna uma resposta em relação a escolha do modelo, foi aplicado o Teste da razão de verossimilhança entre os modelos gerais da distribuição Exponencial e Weibull, que tiveram performance parecida nos métodos de análise gráfica. O teste retornou um p-valor $< 0,0001$ e confirmou a utilização do modelo mais geral (Weibull).

Com a distribuição de probabilidade escolhida, é realizada a identificação e estudo da relação entre as covariáveis selecionadas e o evento de inadimplência. Foram construídos todos os modelos possíveis utilizando as quatro covariáveis escolhidas (Apêndice A) e para encontrar o modelo que melhor se ajusta aos dados, foi utilizado o teste da razão de verossimilhança. O modelo escolhido foi

$$\text{Inadimplência} \sim \text{IDADE} + \text{RENDA} + \text{RISCO} + \text{UF}$$

Para avaliar o ajuste global do modelo de regressão Weibull foram construídos os gráficos de Cox-Snell.

Figura 17 – Análise dos resíduos de Cox-Snell do modelo de regressão Weibull ajustado.



Fonte: Elaboração própria

A análise dos resíduos de Cox-Snell foi satisfatória para o modelo ajustado (figura 17), principalmente quando comparado a função de risco acumulado dos resíduos calculados segundo o modelo ajustado, a curva de sobrevivência dos resíduos manteve-se alinhada aos resíduos de Kaplan-Meier nos momentos iniciais.

O quadro 1 apresenta o valor dos coeficientes encontrados no modelo paramétrico aplicado, todas as variáveis listadas retornaram um Erro-Padrão abaixo do coeficiente estimado e um p-valor $<0,05$, ou seja, podemos rejeitar a hipótese nula de que o coeficiente é igual a zero e assumir que as estimativas encontradas representam o peso no tempo até a inadimplência dos clientes da empresa A. O modelo ajustado retornou coeficientes menor que zero para os clientes com classificação Baixo Risco (Modelo Aquisição), Médio risco (Modelo Aquisição), Alto Risco (Modelo Aquisição), UF-MG e UF-Demais, já as variáveis de Idade e Renda tem influência pequena sobre o tempo de inadimplência dos clientes e a única relação positiva com o tempo de inadimplência está relacionada aos clientes de UF-SE.

Grande parte das variáveis selecionadas no modelo tem estimativas negativas, indicando uma relação inversa com o tempo de inadimplência dos clientes. A variável relacionada a UF dos clientes tem maior discrepância entre as estimativas encontradas, esse processo pode estar ligado ao reconhecimento/credibilidade da marca nos diferentes estados.

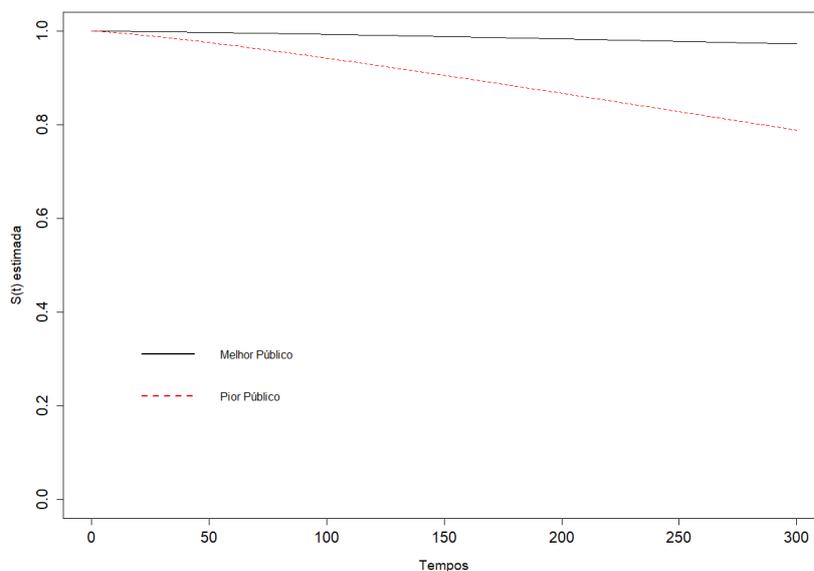
Quadro 1 – Estimativa dos parâmetros do modelo de regressão Weibull

Covariável	Estimativa	Erro-Padrão	Valor p
(Intercept)	10,348	0,075	<0,0001
Idade	-0,002	0,000	0,0069
RENDA	-0,0002	0,000	<0,0001
Baixo Risco (Modelo Aquisição)	-2,154	0,063	<0,0001
Médio Risco (Modelo Aquisição)	-2,688	0,057	<0,0001
Alto Risco (Modelo Aquisição)	-3,854	0,056	<0,0001
UF-Demais	-0,636	0,035	<0,0001
UF-MG	-1,815	0,054	<0,0001
UF-SE	0,350	0,036	<0,0001
Log(scale)	0,306	0,008	<0,0001

Fonte: Elaboração própria

Foram selecionados os clientes que tem UF = SE e classificação de risco = Baixo ou Alto, com o objetivo de observar o pior e o melhor resultado possível, respectivamente, para os cliente de Sergipe, sem levar em consideração as covariáveis Idade e Renda, devido a seu baixo impacto no tempo de inadimplência segundo o modelo. A figura 18 mostra as curvas estimadas utilizando o modelo de Weibull ajustado, evidenciando a diferença de sobrevida logo após as primeiras semanas e uma sobrevida próxima a 95% para os clientes com classificação de Baixo Risco, evidenciando uma boa discriminação da faixa de risco da aquisição no processo de tempo de inadimplência.

Figura 18 – Sobrevivência do grupo com a menor e maior sobrevida possível com base no modelo de regressão Weibull ajustado.



Fonte: Elaboração própria

7 CONCLUSÃO

O cálculo da probabilidade de inadimplência, medido através do escore e obtido a partir de um modelo, é uma técnica amplamente utilizada no mercado de crédito. Porém a rápida evolução do setor e o aumento da concorrência enfatizaram o uso de outras abordagens, que tem o objetivo de auxiliar a tomada de decisão mais rápida e assertiva possível.

De maneira a suprir essa demanda, a análise de sobrevivência vem sendo empregada no mercado de crédito por ser capaz de viabilizar o estudo dos clientes com base no tempo até o evento de interesse, servindo como análise complementar aos modelos já aplicados. Nesse contexto, foi empregado o modelo paramétrico Weibull, com o objetivo de entender a influência das covariáveis de perfil e a classificação do modelo da aquisição no tempo até a inadimplência dos clientes de uma empresa.

A análise indicou um ajuste satisfatório do modelo, com as covariáveis Faixa de risco da aquisição e Unidade Federativa representando a maior influência sobre o tempo até inadimplência dos clientes. Levando esses fatores em consideração, a empresa A pode realizar ações de contenção de inadimplência com base no desempenho das covariáveis apresentadas no modelo.

Um fator que não foi discutido no trabalho, é a criação de um escore de crédito que ajude a relacionar o modelo comportamental que utiliza regressão logística e o modelo de análise de sobrevivência aplicado. Um ponto de melhoria seria incluir mais covariáveis ao estudo, principalmente as que levem em consideração os hábitos de consumo dos clientes, o que não foi possível no presente estudo devido a restrição de algumas variáveis do banco de dados por parte da empresa A. O teste com outras distribuições, pode melhorar os resultados apresentados.

REFERÊNCIAS

- ABECS, A. B. das Empresas de Cartões de Crédito e S. *User Manual for glossaries.sty*. [S.l.], 2022. Disponível em: <<https://www.abecs.org.br/apresentacoes-e-estudos/Apresentação-Coletiva-1S22-2T22.pdf>>. Acesso em: 25 set. 2022. Citado na página 14.
- ABREU, E. G. de. *Sistema Financeiro Nacional*. São Paulo, Brasil: MÉTODO, 2017. Citado 2 vezes nas páginas 10 e 14.
- ANDERSON, R. *The Credit Scoring Toolkit, Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York, United States: Oxford University, 2007. Citado na página 16.
- ARAÚJO, E. A. *Modelagem de risco de crédito: aplicação de modelos Credit Scoring no Fundo Rotativo de Ação da Cidadania – Cred Cidadania*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Pernambuco, 2006. Citado 2 vezes nas páginas 10 e 13.
- AVERY, R. B.; BREVOORT, K. P.; CANNER, G. B. Credit scoring and its effects on the availability and affordability of credit. *The Journal of Consumer Affairs*, Wiley, v. 43, n. 3, p. 516–530, 2009. ISSN 00220078, 17456606. Disponível em: <<http://www.jstor.org/stable/23859699>>. Citado na página 15.
- BORGES, A. I. M. *Análise de Sobrevivência com o R*. Dissertação (Dissertação de Mestrado) — Universidade da Madeira, 2014. Citado na página 19.
- CAMINHA, L. *Novo Mercado de Crédito*. São Paulo, Brasil: Revista dos Tribunais; Nova Edição^a (3 maio 2021), 2021. Citado na página 14.
- CAO, R.; VILAR, J.; RIVERA, A. E. D. Modelling consumer credit risk via survival analysis. *Sort: Statistics and Operations Research Transactions*, ISSN 1696-2281, Vol. 33, Nº. 1, 2009, pags. 3-30, v. 33, 08 2009. Citado 2 vezes nas páginas 13 e 21.
- CARVALHO, M. S. et al. *Análise de Sobrevivência: teoria e aplicação em saúde*. Rio de Janeiro, Brasil: Estação FIOCRUZ, 2011. Citado 4 vezes nas páginas 21, 22, 23 e 24.
- CHAIA, A. J. *MODELOS DE GESTÃO DO RISCO DE CRÉDITO E SUA APLICABILIDADE AO MERCADO BRASILEIRO*. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo, 2003. Citado 2 vezes nas páginas 13 e 15.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. Rio de Janeiro, Brasil: Blucher, 2006. Citado 7 vezes nas páginas 17, 19, 20, 21, 22, 24 e 25.
- COSTA, F. N. da; COSTA, C. A. N. da; OLIVEIRA, G. C. de. *Mercado de cartões de pagamentos no Brasil*. São Paulo, Brasil: ABECS -- ASSOCIAÇÃO BRASILEIRA DAS EMPRESAS DE CARTÕES DE CRÉDITO E SERVIÇOS, 2010. Citado na página 14.
- DANTAS, D. H. C. *Mercados de Crédito e de Capitais*. São Paulo, Brasil: Editora Senac São Paulo, 2020. Citado na página 15.

- DEMIRGUC-KUNT, A.; KLAPPER, L.; SINGER, D. Financial inclusion and inclusive growth : A review of recent empirical evidence. No. 8040. World Bank, Washington, DC. © World Bank, 2017. Disponível em: <<https://openknowledge.worldbank.org/handle/10986/26479>>. Citado na página 16.
- FERNANDES, G. B.; ARTES, R. *Spatial correlation in credit risk and its improvement in credit scoring*. [S.l.], 2013. Disponível em: <https://ideas.repec.org/p/ibm/ibmecp/wpe_321.html>. Citado na página 15.
- HSIEH, H.-I.; LEE, T.-P.; LEE, T.-S. Data mining in building behavioral scoring models. In: . [S.l.: s.n.], 2011. p. 1 – 4. Citado 2 vezes nas páginas 16 e 17.
- LI, X.-L.; ZHONG, Y. An overview of personal credit scoring: Techniques and future work. *International Journal of Intelligence Science*, v. 02, p. 181–189, 01 2012. Citado na página 17.
- MAFIO, F. *Risco Nosso De Cada Dia - Uma Orientacao Objetiva Para Profissionais*. São Paulo, Brasil: Estação das Letras Editora, 2007. Citado na página 18.
- MANOS, M. G. L. *Sistema Financeiro e Desenvolvimento Econômico: O Papel do Microcrédito*. Aracaju, Brasil: Embrapa, 2006. Citado 2 vezes nas páginas 10 e 13.
- MOORE, D. F. *Applied Survival Analysis Using*. Nova Jersey, EUA: Springer, 2016. Citado 2 vezes nas páginas 19 e 23.
- NARAIN, B. Survival analysis and the credit granting decision. *Credit Scoring and Credit Control*, p. 109 – 201, 1992. Citado na página 17.
- PRESENTE, R. *Mercados Financeiros*. Salvador, Bahia, Brasil: UFBA, 2019. Citado na página 10.
- SAUDERS, A. *Medindo o Risco de Crédito*. Rio de Janeiro, Brasil: Qualitymark, 2000. Citado na página 25.
- SCHRICKEL, W. K. Análise de crédito: concessão e gerência de empréstimos. *Editora Atlas S.A. 2000*, v. 4, 2000. Citado na página 14.
- SEBBEN, N. A. *MERCADO DE CRÉDITO, LGPD E CADASTRO POSITIVO: REFLEXÕES ACERCA DA LEI Nº 13.709/2018 E DA LEI COMPLEMENTAR Nº 166/2019*. Londrina, Brasil: Editora Thoth, 2021. Citado 2 vezes nas páginas 15 e 16.
- SHANG, H.; SONG, Q.; WU, Y. Credit market development and firm innovation. *ADB Working Paper Series*, No. 649, 01 2017. Citado na página 14.
- SOUZA, M. A. P. de. *O uso do crédito pelo consumidor: percepções multifacetadas de um fenômeno intertemporal*. Dissertação (Dissertação de Mestrado) — Universidade de Brasília, 2013. Citado na página 15.
- STEPANOVA, M.; THOMAS, L. Survival analysis methods for personal loan data. *Operations Research, INFORMS*, v. 50, n. 2, p. 277–289, 2002. ISSN 0030364X, 15265463. Disponível em: <<http://www.jstor.org/stable/3088495>>. Citado na página 17.
- THOMAS, L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, v. 16, n. 2, p. 149–172, 2000. ISSN 0169-2070. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207000000340>>. Citado na página 15.

APÊNDICE A – Modelos

Modelos

Inadimplência ~ 1

Inadimplência ~ IDADE

Inadimplência ~ RENDA

Inadimplência ~ C. RISCO

Inadimplência ~ UF

Inadimplência ~ IDADE + RENDA

Inadimplência ~ IDADE + C. RISCO

Inadimplência ~ IDADE + UF

Inadimplência ~ RENDA + C. RISCO

Inadimplência ~ RENDA + UF

Inadimplência ~ C. RISCO + UF

Inadimplência ~ IDADE + RENDA + C. RISCO

Inadimplência ~ IDADE + UF + C. RISCO

Inadimplência ~ RENDA + UF + C. RISCO

Inadimplência ~ IDADE + RENDA + UF

Inadimplência ~ IDADE + RENDA + UF + C. RISCO
