



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

Arnaldo Dantas Barreto Neto

Uso da análise de sobrevivência na avaliação de ativos contratuais

São Cristóvão – SE

2025

Arnaldo Dantas Barreto Neto

Uso da análise de sobrevivência na avaliação de ativos contratuais

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

2025

São Cristóvão – SE2025

Arnaldo Dantas Barreto Neto

Uso da análise de sobrevivência na avaliação de ativos contratuais

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.

Aprovado em ____/____/____, Nota Final ____.

Banca Examinadora:

Prof. Dr. Cleber Martins Xavier
Orientador

Profa. Dra. Eucymara Franca Nunes Santos
1º Examinadora

Profa. Dra. Cristiane Silva Corrêa
2º Examinadora

AGRADECIMENTOS

Ao Deus Todo Poderoso, que me concedeu saúde, disposição e esperança em um mundo melhor.

À minha esposa Izildinha o apoio e compreensão pelas ausências durante esse período de graduação.

Aos meus pais: Murilo (*in memoriam*) e Creusa que sempre me ressaltaram as virtudes da curiosidade, do esforço e da renúncia.

À minha família pelos estímulos e gestos de afeto nessa decisão de cursar ciências atuariais.

Ao meu orientador Prof. Cleber Xavier de sabedoria e paciência inesgotáveis, pelo aceite da minha proposta, direcionamentos imprescindíveis e disponibilidade franca.

Aos colegas pelo companheirismo, incentivo e cooperação.

Aos professores e funcionários do DECAT pela atenção e dedicação durante todo o curso.

"Somos peregrinos nessa terra...Não sabemos até quando! Devemos encarar a Vida...Não com tristeza, mas com seriedade e esperança."

S. João Paulo II

RESUMO

Investigou-se a aplicação de análise de sobrevivência na avaliação de ativos contratuais no contexto de correspondentes bancários (CORBAN), destacando a relevância dessa técnica para modelar e prever fluxos financeiros futuros, garantindo a sustentabilidade econômica dessas entidades. Utilizando uma ampla base de dados de contratos de empréstimos consignados e o software RStudio, foram explorados modelos paramétricos, como Weibull e log-normal, e semiparamétricos, como o modelo de riscos proporcionais de Cox, para analisar a duração dos contratos e identificar variáveis determinantes. Os resultados evidenciaram que a maior parte das falhas ocorre nos primeiros 24 meses, com modelos Weibull apresentando bom ajuste para descrever a função de sobrevivência, enquanto o modelo de Cox confirmou a influência significativa das covariáveis banco financiador, tipo do contrato, linha de empréstimo e grupo. Observou-se que a sobrevivência dos contratos varia de acordo com essas características, fornecendo insights importantes para o planejamento financeiro e estratégias de pricing. A pesquisa também destaca a necessidade de ferramentas estatísticas para apoiar a gestão econômica e financeira de ativos contratuais. A análise de sobrevivência mostrou-se eficaz ao fornecer bases quantitativas para tomadas de decisão e propõe refinamentos futuros para incorporar mudanças dinâmicas nos contratos e suas variáveis.

Palavras Chave: Análise de Sobrevivência, Ativos contratuais, Correspondentes bancários (CORBAN), Fluxos financeiros, Modelos paramétricos, Modelo de Cox, Covariáveis

ABSTRACT

The application of survival analysis in the evaluation of contractual assets within the context of banking correspondents (CORBAN) was investigated, highlighting the relevance of this technique to model and forecast future financial flows, ensuring the economic sustainability of these entities. Using a comprehensive database of payroll loan contracts and the RStudio software, parametric models such as Weibull and log-normal, as well as semi-parametric models like the Cox proportional hazards model, were explored to analyze contract duration and identify determinant variables. The results showed that most failures occur within the first 24 months, with Weibull models presenting a good fit to describe the survival function, while the Cox model confirmed the significant influence of covariates such as financing bank, contract type, loan line, and group. It was observed that contract survival varies according to these characteristics, providing important insights for financial planning and pricing strategies. The research also underscores the need for statistical tools to support the economic and financial management of contractual assets. Survival analysis proved effective in providing quantitative bases for decision-making and suggests future refinements to incorporate dynamic changes in contracts and their variables.

Keywords: Survival Analysis, Contractual Assets, Banking Correspondents (CORBAN), Financial Flows, Parametric Models, Cox Model, Covariates.

LISTA DE ILUSTRAÇÃO

Figura 1 – Percentagem de censuras e falhas presentes no banco de dados cedidos pelo CORBAN.	20
Figura 2 – Distribuição do nº de falhas quanto ao tempo de permanência em meses	20
Figura 3 – Estimativa Kaplan-Meier para o tempo sobrevivência dos contratos	21
Figura 4 – EKM e log-rank estratificado para a covariável família	22
Figura 5 – EKM e log-rank estratificado para a covariável tipo proposta	22
Figura 6 – EKM e log-rank estratificado para a covariável linha empréstimo	23
Figura 7– EKM e log-rank estratificado para a covariável Banco	23
Figura 8– Modelos probabilísticos comparados com o EKM	24
Figura 9– Ajuste dos modelos probabilísticos pelos métodos gráficos	24
Figura 10 – curvas de sobrevivência estimada pelos modelos Weibull e EKM	25
Figura 11 – Resíduos Cox-Snell versus tempo	26
Figura 12 – Resíduos deviance	26
Figura 13 – Curva de sobrevivência e Taxa de risco para Bancos 2 e 5	28
Figura 14 – Coeficientes de correlação de Pearson entre os resíduos de Schoenfeld	29
Figura 15 – Coeficientes de correlação de Pearson entre os resíduos de Schoenfeld padronizados	29
Figura 16 – Resíduos score-dfbeta	30
Figura 17 – Resíduos deviance	30
Figura 18 – Razão de riscos para as covariáveis	32
Figura 19 – Curvas de sobrevivência para família INSS e CONVENIO	32

LISTA DE TABELAS

Tabela1 - Comparativo Ativos Contratuais versus Recebíveis	5
Tabela 2 - Descrição das variáveis utilizadas e fornecidas pelo CORBAN	9
Tabela 3 – Categorias, funções chave e propósitos do pacote survival	10
Tabela 4 – Categorias, funções chave e propósitos do pacote survminer	10
Tabela 5 – Comparativo entre os principais modelos paramétricos	15
Tabela 6 - Valores do AIC para os modelos probabilísticos	25
Tabela 7 - Estimativa dos parâmetros do modelo de regressão Weibull	27
Tabela 8 – Estimativas dos coeficientes do modelo Cox	31

LISTA DE ABREVIATURAS E SIGLAS

CORBAN – Correspondente bancário

AS – Análise de Sobrevivência

EKM – Estimador Kaplan – Meier

TRV – Teste da razão de verossimilhança

AIC- Critério de informação AKAIKE

SUMÁRIO

1 INTRODUÇÃO	1
2 OBJETIVOS	2
2.1 Geral	2
2.2 Específicos	2
3 JUSTIFICATIVA	3
4 REVISÃO LITERÁRIA	4
4.1 Ativos Contratuais e Recebíveis	4
4.2 Análise de sobrevivência e contratos de serviços	6
5 METODOLOGIA	6
5.1 Material	9
5.2 Método	11
6 RESULTADOS	20
6.1 Análise Descritiva e Exploratória	20
6.2 Estimador de Kaplan- Meier	21
6.3 Modelagem paramétrica	24
6.4 Modelagem semi-paramétrica	28
7 CONCLUSÃO	33
REFERÊNCIAS BIBLIOGRÁFICAS	

1 INTRODUÇÃO

O mercado de empréstimos consignados a aposentados e trabalhadores do setor público vem crescendo ao longo dos anos refletindo o crescimento demográfico dessas categorias. Segundo o BACEN, em 2010 o saldo da carteira de empréstimos apontava R\$ 41 bilhões e R\$ 83 bilhões, atingindo em 2024 os montantes de R\$ 277 bilhões e R\$370 bilhões respectivamente e mais de 18 milhões de mutuários, situando-se no topo dos ativos bancários, daí a importância socioeconômica desse tipo de instrumento financeiro. Buscando aumentar a capilaridade junto à população, as instituições financeiras recorrem aos correspondentes bancários (CORBAN), entidades empresariais que fazem a captação de tomadores de empréstimos para o banco financiador.

A Resolução Nº 4.294 do Conselho Monetário Nacional, de 20/12/2013 estabelece que a remuneração por esses serviços deve ser feita em duas etapas: uma parte no ato da contratação e outra em parcelas mensais que são pagas condicionalmente caso o contrato permaneça ativo junto a instituição financeira. Entre as causas de inatividade destacam-se a portabilidade para outro financiador ou a liquidação antecipada, entre outras variações contratuais.

A mensuração da receita total a valor presente e dos fluxos futuros de um determinado contrato é fundamental para o planejamento econômico-financeiro do CORBAN. A viabilização do negócio envolve custos de captação da ordem de setenta por cento da receita operacional líquida auferida por um CORBAN (notas do Autor), pagos à vista, que são parcialmente cobertos mediante empréstimos a serem amortizados com a própria receita das comissões auferidas em recorrência mensal. Este estudo apontou que 50% dos contratos se tornam inativos até os 46 meses, interrompendo o fluxo de receitas, é importante investigar as variáveis envolvidas nessa dinâmica, mensurar os seus efeitos para assegurar a solvência e liquidez da entidade empresarial.

O uso das técnicas de análise de sobrevivência (AS) parece promissora para determinar valores estatísticos capazes de fazer avaliações e previsões a esse respeito. Assim, foi utilizado um banco de dados cedido por um CORBAN com larga e ampla história no mercado e aplicados modelos de regressão paramétricos e semiparamétrico de Cox. A ideia central é descrever as funções de risco e sobrevivência para, a partir delas, desenvolver aplicações ligadas ao *pricing*, *valuation*, orçamentação, planejamento financeiro. Trata-se de um estudo inicial que pode levar a refinamentos capazes de enfrentar situações práticas mais complexas.

2 OBJETIVOS

2.1 Geral

Estudar metodologicamente as capacidades dos modelos de regressão paramétricos e semiparamétrico de Cox na investigação de fluxos financeiros de componentes dos ativos contratuais sob risco em CORBANs.

2.2 Específicos

- Descrever a sobrevida dos contratos utilizando o estimador de Kaplan-Meier;
- Ajustar modelos de regressão paramétricos e modelos de Cox;
- Verificar a significância estatística de variáveis importantes aos contratos;
- Verificar a adequação dos modelos de regressão;
- Estimar curvas de sobrevivência e risco utilizando os modelos estimados.

3 JUSTIFICATIVA

Existe uma lacuna na avaliação e apuração de resultados de entidades empresariais em que a presença de ativos contratuais sob risco compõe parte significativa do balanço patrimonial e das movimentações com repercussão econômica. Especificamente para os CORBANs, essa categoria de direitos representam o principal lastro a garantir a solvência e liquidez de empréstimos junto a instituições financeiras. Todavia, gestores financeiros e contadores carecem de ferramentas estatísticas e atuariais capazes de realizar essas tarefas.

Segundo a FEBRABAN, em 2019 existiam 274 mil CORBANs ativos na oferta de crédito. Todavia, após o impulso de digitalização dos negócios pós pandemia, esse segmento de negócios encontra-se em processo de consolidação com redução do nº de *players*. A má avaliação dos riscos sobre os ativos constitui o principal motivo para o enfraquecimento econômico dessas entidades, resultando no encerramento de suas atividades ou na absorção por outras mais bem posicionadas.

A análise de sobrevivência pode ser uma técnica importante para atender essa necessidade, propiciando mais assertividade quanto à constituição de provisões e reservas, *valuation, pricing*, orçamentação e planejamento empresarial. A verificação da adequação das modelagens paramétricas e semiparamétricas podem trazer novos horizontes e melhor compreensão a respeito das variáveis que impactam na durabilidade dos contratos e, conseqüentemente, no seu valor.

4 REVISÃO LITERÁRIA

4.1 Ativos Contratuais e Recebíveis

Os Ativos Contratuais são uma relevante categoria de direitos no campo empresarial, em alguns casos podem representar a maior parte do Ativo total. A norma IFRS 15 - Receita de Contratos com Clientes estabelece um modelo abrangente para o reconhecimento de receita oriunda de contratos com clientes, incluindo a identificação e mensuração de ativos contratuais. De acordo com Deloitte (2020), um ativo contratual surge quando uma entidade transfere bens ou serviços a um cliente antes de receber a contrapartida ou antes que o pagamento seja devido, desde que o direito ao pagamento não seja condicionado apenas à passagem do tempo. Essencialmente, os ativos contratuais representam o direito da entidade de receber a contraprestação em troca dos bens ou serviços transferidos, mas essa contraprestação ainda não é incondicional, ou seja, depende da continuidade do contrato ou de algum evento futuro.

Esse conceito é fundamental para empresas que operam com contratos de serviços de longo prazo, como telecomunicações, ou prestação de serviços financeiros. PwC (2019) destaca que o IFRS 15 exige a identificação de obrigações de desempenho distintas dentro de um contrato, e os ativos contratuais são reconhecidos quando essas obrigações são parcialmente satisfeitas. Já EY (2021) enfatiza a importância da mensuração desses ativos com base no valor presente dos fluxos de caixa esperados, considerando descontos, inadimplência e outros fatores de risco.

Diferenciam-se dos recebíveis, definidos pela IFRS 9 – Instrumentos Financeiros. Recebíveis que representam direitos contratuais a receber, caixa ou outro ativo financeiro de outra entidade. Esses direitos são incondicionais, ou seja, a empresa já cumpriu a sua parte do contrato, e o recebimento do valor é garantido, sujeito apenas ao risco de crédito.

Uma síntese comparativa incluindo: a definição, a norma IFRS aplicável e a condicionalidade para esses dois tipos de ativos estão na Tabela 1:

Tabela1 - Comparativo Ativos Contratuais versus Recebíveis

Característica	Ativos Contratuais	Recebíveis
Definição	Direito de receber contraprestação, condicionado ao contrato	Direito incondicional de receber caixa
Norma IFRS	IFRS 15 – Receita de Contratos com Clientes	IFRS 9 – Instrumentos Financeiros
Condição	Condicionado a eventos futuros ou à continuidade do contrato	Incondicional, sujeito apenas ao risco de crédito

Fonte: Elaborado pelo autor

No contexto brasileiro, Souza e Lima (2022) exploram a adaptação do IFRS 15 às práticas contábeis locais, destacando desafios na classificação de ativos contratuais em setores regulados, como energia e seguros. Por fim, KPMG (2018) aponta que a norma exige divulgações detalhadas sobre ativos contratuais, como sua natureza, valor e mudanças ao longo do tempo, promovendo transparência para usuários das demonstrações financeiras. Esses princípios são particularmente relevantes para análises atuariais, pois conectam a gestão de contratos à previsão de fluxos financeiros futuros.

A principal causa de descontinuidade de um contrato de serviços entre um Correspondente Bancário (CORBAN) e a instituição financeira, cujo objetivo é a captação de clientes tomadores de empréstimos, é a portabilidade do contrato para outra instituição. Introduzida no Brasil em 2006, pela Resolução nº 3.401 do Conselho Monetário Nacional (CMN), mas ganhou maior estrutura com a Resolução nº 4.292, de 20 de dezembro de 2013. Essa norma estabeleceu as diretrizes específicas para a transferência de operações de crédito entre instituições financeiras, incluindo o consignado, que se destaca por sua representatividade no volume de portabilidades efetivadas. A Resolução nº 4.292 foi posteriormente alterada pela Resolução nº 4.762, de 27 de novembro de 2019, que refinou os procedimentos e reforçou a transparência, proibindo custos adicionais ao consumidor na operação.

No caso em estudo, os fluxos financeiros estão condicionados à existência do contrato de empréstimo consignado. A análise de sobrevivência destaca-se como importante ferramenta para o *valuation* desses ativos.

4.2 Análise de sobrevivência e contratos de serviços

A análise de sobrevivência (AS) é uma área da estatística que estuda o tempo até a ocorrência de um evento de interesse, como falhas, cancelamentos ou termos de contratos. Segundo Klein e Moeschberger (2003), trata-se de uma metodologia essencial para modelar dados com censura, ou seja, situações em que o evento não é observado para todos os indivíduos durante o período de estudo. Esse aspecto é particularmente relevante em contratos de serviços, onde nem todos os contratos chegam ao término no período analisado.

Na análise de sobrevivência, o foco está na “função sobrevivência” – que estima a probabilidade de um contrato de serviço permanecer ativo além de um determinado tempo – e na “função risco” ou “*hazard function*”, que expressa a taxa instantânea de ocorrência do evento (rescisão ou falha contratual) em função do tempo. Métodos não paramétricos, como o estimador de Kaplan–Meier (HOSMER et al., 2008; FLEMING & HARRINGTON, 2001), e modelos semiparamétricos, como o modelo de riscos proporcionais de Cox (COX & OAKES, 2004), formam a base dessa abordagem analítica.

Além destes, técnicas paramétricas e modelos para eventos recorrentes também têm sido explorados para captar a complexidade do comportamento temporal dos contratos de serviços (COOK & LAWLESS, 2007; MACHIN, CHEUNG & PARMAR, 2006). Esses métodos possibilitam a incorporação de covariáveis e a identificação de determinantes que influenciam a “vida” de um contrato, contribuindo para uma visão mais acurada dos riscos envolvidos.

Colosimo e Giolo (2006) destacam a importância de técnicas não paramétricas, como o estimador de Kaplan–Meier, e modelos paramétricos, como o de Weibull, para a análise de tempos de sobrevivência. Esses métodos permitem estimar a probabilidade de um contrato permanecer ativo ao longo do tempo. Hosmer, Lemeshow e May (2008) complementam essa visão ao enfatizar o modelo de regressão de Cox, amplamente utilizado para incorporar variáveis explicativas, como características dos clientes ou condições contratuais.

Os modelos estatísticos são fundamentais para a aplicação da análise de sobrevivência em contextos atuariais. Lee e Wang (2003) exploram distribuições probabilísticas, como a exponencial e a log-normal, que são úteis para descrever tempos de falha em serviços. Já Kalbfleisch e Prentice (2002) detalham o uso de modelos semi-

paramétricos, que combinam flexibilidade e robustez, sendo aplicáveis a contratos de serviços com diferentes padrões de término.

A análise de sobrevivência tem sido aplicada em diversos contextos relacionados a contratos de serviços. Segundo Allison (2010), em setores como telecomunicações e seguros, esses métodos ajudam a prever o tempo de retenção de clientes e a estimar riscos de cancelamento. Cox e Oakes (1984), em uma obra revisada e republicada em 2001, destacam como a análise de sobrevivência pode ser adaptada para eventos múltiplos, como renovações ou rescisões parciais de contratos.

O modelo de riscos proporcionais de Cox, proposto por David R. Cox em 1972, é amplamente utilizado em estudos de análise de sobrevivência devido à sua flexibilidade e aplicabilidade em diversas áreas, como medicina, epidemiologia, ciências sociais e econômicas. Este modelo é denominado semiparamétrico por combinar componentes paramétricos, que descrevem os efeitos das covariáveis no risco, e não paramétricos, que representam a função de risco basal de forma livre (Carvalho, S. M. et al, 2011).

Segundo Colosimo e Giolo (2006), o modelo de Cox se baseia na suposição de riscos proporcionais, onde a razão entre os riscos de dois indivíduos é constante ao longo do tempo. Isso significa que o impacto de uma variável explicativa sobre o risco de um evento não varia no decorrer do tempo. Essa suposição é central para a interpretação dos coeficientes estimados, que representam o efeito log-linear das covariáveis sobre o risco.

Uma característica notável do modelo é a separação da estimativa dos coeficientes das covariáveis da função de risco basal. Conforme explicado por Klein e Moeschberger (2003), essa abordagem possibilita a análise da influência das variáveis sem a necessidade de especificar a forma exata da função de risco basal, conferindo ao modelo maior robustez e generalidade.

A estimação dos parâmetros do modelo é realizada por meio do método de máxima verossimilhança parcial, desenvolvido por Cox (1975). Esse método permite estimar os coeficientes das covariáveis sem a necessidade de calcular explicitamente a função de risco basal, o que simplifica a análise e reduz as limitações associadas a suposições paramétricas rigorosas.

No campo atuarial, Dickson, Hardy e Waters (2013) discutem a aplicação de modelos de sobrevivência em cálculos de reservas e precificação de serviços, integrando variáveis como idade do contrato e perfil do cliente. Lawless (2003) reforça a relevância de modelos de confiabilidade para serviços técnicos, como manutenção de equipamentos, onde o tempo até a falha é um indicador-chave.

No Brasil, a análise de sobrevivência tem ganhado relevância em estudos atuariais. Oliveira e Louzada-Neto (2004) investigam sua aplicação em seguros de vida, mas os princípios podem ser estendidos a contratos de serviços. Pinheiro e Nascimento (2016) exploram modelos bayesianos, oferecendo uma perspectiva moderna para lidar com incertezas em dados contratuais. Por fim, Giolo (2014) atualiza as discussões de Colosimo e Giolo (2006), incorporando riscos competitivos e fração de imunes, temas promissores para o contexto brasileiro.

Na aplicação prática, a análise de sobrevivência tem sido utilizada para mensurar a durabilidade e a estabilidade dos contratos de serviços. Trabalhos recentes na área atuarial, como os de Abreu (2019), Santos (2022) e Silva (2021), demonstram a aplicabilidade dessas técnicas para identificar fatores que antecipam a rescisão contratual, permitindo a estimativa de riscos e a proposição de estratégias mitigadoras. Tais estudos enfatizam que a modelagem estatística, aliada à compreensão dos determinantes comportamentais e operacionais, pode melhorar significativamente a gestão dos contratos.

Oliveira (2018) e Costa & Ferreira (2017) destacam, ainda, que o entendimento dos perfis dos usuários e a integração de dados históricos com modelos preditivos são essenciais na identificação de padrões e no desenvolvimento de indicadores de performance para contratos de serviços. Essa combinação de métodos atua como auxílio no gerenciamento dos riscos e na definição de políticas de precificação, que são fundamentais para as seguradoras e administradoras de contratos na área atuarial.

5 METODOLOGIA

5.1 Material

5.1.1 Base de dados

A base de dados utilizada foi cedida por uma empresa atuante como Correspondente Bancário (CORBAN) na captação de clientes para contratação de empréstimos consignados. Compreendem operações realizadas entre os anos 2015 e 2024 em um regime de coortes abertas. Todos os contratos analisados geram um fluxo financeiro mensal enquanto estiverem ativos. A duração do contrato foi definida como o nº de meses entre o seu início e o último fluxo financeiro observado. Os com duração igual a zero foram excluídos da base de dados. O número de registros é superior a 1,7 milhão, mediante uma amostragem aleatória foi reduzido a 100 mil observações. As variáveis analisadas estão relacionadas na Tabela 2.

Tabela 2 - Descrição das variáveis utilizadas e fornecidas pelo CORBAN

Variável	Descrição	Tipo
Codbanco	Código indicativo do banco emissor do contrato	Fator com valores: -2,2,4,5,6,8,999
Nomlinha	Nome da linha do empréstimo	Fator com valores: Cartão Consignado e Empréstimo Consignado
Tiprop	Tipo do contrato	Fator com valores: Compra de Dívida, INDEFINIDO, Novo, Portabilidade Refin.daPortabilidade, Refinanciamento
Família	Família do contrato	Fator com valores: CONVENIOS, INSS OUTROS RPPS
Duration	nº de meses que o contrato esteve ativo	Numérico
Censura	Indica se o dado foi ou não censurado	Numérico (0 = censura;1=falha)

Fonte: Elaborado pelo autor

5.1.2 Suporte Computacional

Utilizou-se o software RStudio 2024.12.0+467, os pacotes requeridos foram: readxl, lubridate, dplyr, tidyr, gridextra, patchwork, survival, survminer e flexsurv.

Mantido por Terry M. Therneau, o pacote survival oferece uma ampla gama de funções, categorizadas em várias áreas-chave, com base na documentação de Survival Analysis version 3.7-0 from CRAN e survival package - RDocumentation. Destacam-se as aplicações nas seguintes áreas: manipulação de dados, estimação não paramétrica,

modelo de Cox e modelos paramétricos. Na Tabela 3 pode-se visualizar as principais funções e suas funcionalidades

Tabela 3 – Categorias, funções chave e propósitos do pacote `survival`

Categoria	Funções Chave	Propósito
Manipulação de Dados	<code>Surv</code> , <code>strata</code> , <code>tcut</code> , <code>tmerge</code>	Criar e manipular dados de sobrevivência
Estimação Não Paramétrica	<code>survfit</code> , <code>plot.survfit</code> , <code>survdiff</code>	Estimar e comparar curvas de sobrevivência
Modelo de Cox	<code>coxph</code> , <code>cox.zph</code> , <code>predict.coxph</code>	Ajustar e diagnosticar o modelo de Cox
Modelos Paramétricos	<code>survreg</code> , <code>predict.survreg</code>	Ajustar e prever a partir de distribuições paramétricas
Modelos de Múltiplos Estados	<code>aareg</code> , <code>finegray</code> , <code>statefig</code>	Modelar transições e riscos competidores
Utilitários	<code>cluster</code> , <code>pspline</code> , <code>brier</code>	Ferramentas adicionais para análise e diagnósticos

Fonte: Elaborado pelo autor

Mantido por Alboukadel Kassambara, o pacote `survminer` está disponível no Comprehensive R Archive Network (CRAN) `survminer: Drawing Survival Curves using 'ggplot2'` e é amplamente utilizado em pesquisas médicas, engenharia e ciências sociais para apresentar resultados de forma clara e visualmente atraente. Destacam-se as aplicações nas seguintes áreas: plotagem de curvas de sobrevivência, diagnósticos de modelos Cox, sumarização de dados. Na Tabela 4 pode-se visualizar as principais funções e funcionalidades.

Tabela 4 – Categorias, funções chave e propósitos do pacote `survminer`

Categoria	Funções Chave	Propósito
Plotagem de Curvas	<code>ggsurvplot</code> , <code>arrange_ggsurvplots</code>	Criar e organizar curvas de sobrevivência
Diagnósticos de Cox	<code>ggcoxzph</code> , <code>ggcoxfunctional</code>	Testar suposições e formas funcionais
Sumarização de Dados	<code>surv_summary</code> , <code>surv_cutpoint</code>	Resumir e categorizar dados de sobrevivência
Utilitários	<code>surv_median</code> , <code>surv_pvalue</code>	Calcular medianas e valores-p

Fonte: Elaborado pelo autor

5.2 Método

5.2.1 Sobrevivência e Risco

O objetivo principal da AS é modelar e estimar a distribuição do tempo de sobrevivência, frequentemente utilizando funções como a função de sobrevivência e a função de risco. Essas funções ajudam a descrever a probabilidade de sobrevivência ao longo do tempo e a taxa instantânea de ocorrência do evento.

Segundo Colosimo e Giolo (2006), a função de sobrevivência, denotada por $S(t)$, é uma função monótona, não crescente e contínua, representa a probabilidade de que o tempo de sobrevivência T (uma variável aleatória) seja maior que um tempo específico t :

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

Em que

T é uma variável aleatória não negativa.

t é um valor específico de tempo.

$S(t)$ varia entre 0 e 1, com $S(0) = 1$ (todos sobrevivem no início) e $S(t) \rightarrow 0$ quando $t \rightarrow \infty$ (eventualmente, todos experimentam o evento, se não houver censura).

A função de sobrevivência está relacionada à função de densidade acumulada $F(t) = P(T \leq t)$ pela equação:

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

Se $f(t)$ é a função de densidade de probabilidade de T , então:

$$f(t) = -\frac{dS(t)}{dt}$$

A função de risco (hazard), ou taxa de falha, denotada por $h(t)$, mede a probabilidade instantânea de o evento ocorrer em t , dado que o indivíduo sobreviveu até esse momento:

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

Diferentemente da função de sobrevivência ele pode assumir diversos formatos. Está diretamente relacionada à função de sobrevivência e à densidade de probabilidade:

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

Substituindo $f(t) = -\frac{dS(t)}{dt}$, temos:

$$h(t) = -\frac{1}{S(t)} \frac{dS(t)}{dt} = -\frac{d}{dt} \ln(S(t))$$

A integral cumulativa da função de risco, chamada de risco acumulado $H(t)$, é definida como:

$$H(t) = \int_0^t h(u) du$$

E está relacionada a $S(t)$ por:

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

Essa relação é fundamental em modelos paramétricos, como o modelo exponencial onde $h(t) = \lambda$, é uma constante ou o modelo de Weibull, em que $h(t) = \lambda p t^{p-1}$, com λ e p sendo parâmetros de escala e forma, respectivamente.

Segundo Souza, J.M.P, 2009, a censura é um conceito central que diferencia essa área de outros métodos estatísticos. Ela ocorre quando o tempo exato do evento de interesse (como morte, falha ou recuperação) não é observado para todos os indivíduos durante o período de estudo. Esse fenômeno é comum em experimentos clínicos, estudos longitudinais ou testes de confiabilidade, e sua presença exige métodos específicos para estimar funções como a de sobrevivência $S(t)$ e a de risco $h(t)$.

A mais frequente é a censura à direita, que acontece quando o evento não ocorre até o final do estudo ou o indivíduo é perdido durante o acompanhamento (e.g., abandona o estudo), é o tipo encontrado nesse estudo.

Caso a data de saída do contrato da carteira de ativos contratuais ocorra antes do prazo final, é interpretada como falha ($\delta = 1$), e a partir dela um determinado contrato deixa de gerar fluxos financeiros para a empresa. Contratos que chegaram até o final ou que permaneciam ativos até a data de 30/06/2024 foram considerados censurados. ($\delta = 0$).

5.2.2 Estimador Kaplan-Meier

O estimador Kaplan-Meier (EKM) descreve a função de sobrevivência $S(t)$, é uma ferramenta não paramétrica, ou seja, não assume uma distribuição específica para os tempos de sobrevivência, o que a torna versátil para diferentes tipos de dados. Essa característica é particularmente útil em contextos em que a distribuição subjacente é desconhecida ou complexa, com sua formulação dada por

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

em que t_i são os tempos distintos em que ocorrem eventos, d_i é o número de eventos (como mortes ou falhas) naquele tempo t_i , e n_i é o número de indivíduos ou itens ainda "em risco" imediatamente antes de t_i . Resulta em uma curva de sobrevivência em degraus, que permanece constante entre os tempos de evento e cai nos momentos em que ocorrem eventos (DUDLEY, W. N et al, 2016)

Os dados censurados são incorporados ao estimador incluindo os indivíduos censurados no número em risco (n_i) até o momento de sua censura, mas sem contar como eventos ($d_i = 0$ nesses pontos).

O EKM é consistente e assintoticamente normal, permitindo a construção de intervalos de confiança para avaliar sua precisão. Sua variância pode ser calculada usando a fórmula de Greenwood:

$$\text{Var}(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

5.2.3 Modelos Probabilísticos - Estimação Paramétrica

A estimação paramétrica é uma abordagem estatística na qual se assume que os tempos de sobrevivência seguem uma distribuição de probabilidade específica, por exemplo: exponencial, Weibull ou log-normal.

5.2.3.1 Modelo Exponencial: Taxas de Risco Constantes

O modelo exponencial é o mais simples, caracterizado por uma taxa de risco constante no tempo, habitualmente é referida como uma "distribuição sem memória". A função de sobrevivência e a função de risco são dadas por (Colosimo, E.A. e Giolo, S.R., 2006):

$$\text{Função de Sobrevida: } S(t) = e^{-\lambda t}$$

$$\text{Função de Risco: } h(t) = \lambda$$

Para a estimação de parâmetros com dados censurados à direita, a estimativa de máxima verossimilhança de λ é:

$$\hat{\lambda} = \frac{d}{\sum t_i}$$

onde d é o número de eventos observados, e $\sum t_i$ é a soma de todos os tempos observados.

5.2.3.2 Modelo Weibull: Taxas de Risco Flexíveis

O modelo de Weibull oferece maior flexibilidade, permitindo taxas de risco crescentes, decrescentes ou constantes, dependendo de seu parâmetro de forma k . Suas funções de sobrevivência e de risco são:

$$\text{Função de Sobrevida: } S(t) = e^{-(t/\sigma)^k}$$

$$\text{Função de Risco: } h(t) = \frac{k}{\sigma} \left(\frac{t}{\sigma}\right)^{k-1}$$

Aqui, σ é o parâmetro de escala, e k determina a forma:

Se $k = 1$, reduz-se ao modelo exponencial (taxa de risco constante).

Se $k > 1$, a taxa de risco aumenta com o tempo (por exemplo, componentes envelhecendo).

Se $k < 1$, a taxa de risco diminui com o tempo (por exemplo, falhas iniciais em novos sistemas).

A estimação de parâmetros envolve maximizar a função de verossimilhança, que para dados censurados é:

$$L = \prod_{i=1}^n \left[\frac{k}{\sigma} \left(\frac{t_i}{\sigma}\right)^{k-1} e^{-(t_i/\sigma)^k} \right]^{\delta_i} \left[e^{-(t_i/\sigma)^k} \right]^{1-\delta_i}$$

O modelo de Weibull é amplamente utilizado em engenharia de confiabilidade para modelar tempos de falha e em pesquisas médicas para análise de sobrevivência, como estudos de câncer (NIIYAMA, A.C, 2013).

5.2.3.3 Modelo Log-Normal: Dados com Assimetria Positiva

O modelo log-normal assume que o logaritmo do tempo de sobrevivência segue uma distribuição normal, tornando-o adequado para dados com assimetria positiva, comum em contextos biológicos e econômicos. Sua função de sobrevivência é:

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$$

em que Φ é a função de distribuição cumulativa da distribuição normal padrão, e μ e σ são a média e o desvio padrão de $\ln T$, respectivamente. A função de risco para a distribuição log-normal não tem uma expressão fechada simples, o que pode complicar a interpretação comparada aos modelos exponencial e Weibull. A função de verossimilhança para dados censurados é:

$$L = \prod_{i=1}^n \left[\frac{1}{t_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln t_i - \mu)^2}{2\sigma^2}} \right]^{\delta_i} \left[1 - \Phi \left(\frac{\ln t_i - \mu}{\sigma} \right) \right]^{1-\delta_i}$$

A estimação de parâmetros novamente envolve maximizar esta verossimilhança, frequentemente de forma numérica, devido à presença da função CDF normal na parte censurada. Muito útil quando se tem tempos de sobrevivência com assimetria positiva.

Na Tabela 5 pode-se ver um comparativo entre os principais modelos, o exponencial com apenas um parâmetro, uma vez que a taxa de risco é constante, já os modelos Weibull e log-normal além do parâmetro de escala possuem o de forma.

Tabela 5 – Comparativo entre os principais modelos paramétricos

Modelo	Função de Sobrevida	Função de Risco	Parâmetros Chave	Adequação
Exponencial	$S(t) = e^{-\lambda t}$	$h(t) = \lambda$	λ	Taxa de risco constante, sem memória
Weibull	$S(t) = e^{-(t/\sigma)^k}$	$h(t) = \frac{k}{\sigma} (t/\sigma)^{k-1}$	k, σ	Taxa de risco variável (aumentando/diminuindo)
Log-Normal	$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$	Sem expressão fechada simples	μ, σ	Dados com assimetria positiva

Fonte: Elaborado pelo autor

5.2.3.4 Função Gama Generalizada

Trata-se de uma função que inclui como casos especiais a distribuição exponencial e a Weibull, com o desenvolvimento do pacote flexsurv no R tornou-se possível ajustar uma função gama generalizada aos dados estimando-se os três parâmetros que a caracterizam. Serve como base de comparação do ajuste dos modelos mais simples.

5.2.4 Modelos de Regressão

Modelos de regressão na análise de sobrevivência são usados para estudar como variáveis (covariáveis) afetam o tempo até um evento, como morte ou falha de um

sistema. Eles lidam com dados censurados e incluem tantos modelos paramétricos (ex.: exponencial, Weibull) quanto os semi-paramétricos, a exemplo do modelo de Cox.

5.2.4.1 Ajuste e escolha dos modelos paramétricos

O ajuste de um modelo paramétrico aos dados se inicia com o cálculo dos parâmetros específicos para cada curva. No software R, utiliza-se a função *survreg* do pacote *survival* para a estimação dos parâmetros das três distribuições estudadas. As curvas parametrizadas e a curva do EKM são obtidas pelo comando *ggsurvplot*.

A escolha do melhor modelo paramétrico é realizada por meio de dois conjuntos de métodos: gráficos e estatísticos (teste da razão de verossimilhança e o critério de informação Akaike – AIC).

Os métodos gráficos têm como fundamento a comparação entre as curvas de sobrevivência $S(t)$ do modelo e a obtida pelo EKM, caso haja concordância perfeita os pontos estarão da diagonal $y=x$. (Método 1) ou a linearização por meio de transformação adequada para o modelo (Método 2), quanto mais se aproximar da reta, melhor o ajuste.

Já os métodos estatísticos comparam os modelos mais simples com a gama generalizada e aplicam um teste de hipóteses onde H_0 = o modelo de interesse é adequado versus uma hipótese alternativa de que o modelo não é adequado. O teste da razão de verossimilhança (TRV) cuja estatística de teste tem uma distribuição qui-quadrado com graus de liberdade igual a diferença de parâmetros entre a função gama generalizada e o modelo em teste.

O teste baseado no Critério de Informação de Akaike (AIC), determina como melhor modelo aquele com o menor valor de AIC. Ele penaliza a adição de novas covariáveis, desencorajando overfitting, ou seja, a seleção de um modelo extremamente complexo e com muitos parâmetros que tenham um pobre desempenho preditivo. Deve ser usado suplementarmente ao TRV ou quando esse último for inconclusivo.

5.2.4.2 Análise das covariáveis categóricas

Na modelagem de curvas de sobrevivência, quando se tem covariáveis categóricas, é necessário verificar se as categorias devem ser incluídas no modelo ou se podem ser consideradas como única categoria. A verificação de se pelo menos uma das categorias causa diferenças na curva de sobrevivência foi realizada mediante a aplicação do teste log-rank.

O teste log-rank compara grupos definidos por variáveis qualitativas, ou seja, categóricas, como tratamento (ex.: grupo A vs. grupo B), gênero (masculino vs. feminino) ou nível de exposição (baixo vs. alto). Ele testa a hipótese nula de que não há diferença entre as populações nos tempos de sobrevivência, ou seja, $H_0: S_1(t) = S_2(t)$ para dois grupos, onde $S_1(t)$ e $S_2(t)$ são as funções de sobrevivência de cada grupo. A estatística do teste é baseada na comparação entre o número de eventos observados em cada grupo e o número esperado sob a hipótese nula, considerando todos os tempos de evento distintos em ordem crescente, ajustando para indivíduos censurados. Para cada uma das covariáveis categóricas foram desenvolvidas curvas de sobrevivência pelo EKM como preconizado por Colosimo, E.A. e GIOLO, S.R. (2006).

5.2.4.3 Modelo de regressão paramétrico

Escolhida a função representativa, sem considerar covariáveis, busca-se encontrar um modelo o mais simples possível que explique os dados analisados.

Modelo Exponencial: Assume uma taxa de risco constante: $h(t | X) = \lambda e^{\beta X}$

$S(t | X) = e^{-\lambda t e^{\beta X}}$ onde λ é a taxa basal (risco sem covariáveis), e βX é o efeito linear das covariáveis X .

Modelo Weibull: Permite taxas de risco variáveis: $h(t | X) = k \lambda t^{k-1} e^{\beta X}$

$S(t | X) = e^{-(\lambda t)^k e^{\beta X}}$ k é o parâmetro de forma (risco aumenta se $k > 1$, diminui se $k < 1$), e λ é a escala.

Modelo Log-Normal: Assume que $\ln T$ é normal, com sobrevivência: $S(t | X) = 1 - \Phi\left(\frac{\ln t - (\mu + \beta X)}{\sigma}\right)$ Onde μ é a média basal, e σ é o desvio padrão.

Os coeficientes β quantificam o impacto das covariáveis na taxa de risco ou no tempo de sobrevivência:

No **Modelo Exponencial:** e^{β} é o multiplicador do risco por unidade de X . Exemplo: se $\beta = -0,3$ para "tratamento = 1", o risco é $e^{-0,3} \approx 0,74$ vezes menor para tratados, sugerindo benefício.

O tempo médio de sobrevivência é $1/(\lambda e^{\beta X})$, então β inversamente afeta o tempo esperado.

No **Modelo Weibull:** Similarmente, e^{β} ajusta o risco proporcionalmente. Se $\beta = 0,2$ para "idade", cada ano aumenta o risco em $e^{0,2} \approx 1,22$ vezes.

O parâmetro k interage com β , afetando o tempo esperado ($E[T] = \lambda^{-1}\Gamma(1 + 1/k)e^{-\beta/k}$), tornando a interpretação mais complexa.

No **Modelo Log-Normal**: β desloca a média logarítmica ($\mu + \beta X$), afetando a mediana do tempo de sobrevivência ($e^{\mu + \beta X}$). Um $\beta > 0$ aumenta o tempo mediano.

5.4.2.4 Modelo de regressão semiparamétrico

O modelo de Cox é fundamentado na função de risco (hazard function), que representa a probabilidade instantânea de ocorrência do evento em um dado tempo t , dado que o indivíduo sobreviveu até esse ponto. Segundo Colosimo e Giolo (2017) a função de risco condicional às covariáveis X é dada por:

$$h(t | X) = h_0(t)\exp(\beta^T X)$$

em que:

$h_0(t)$ é a função de risco basal, que é a função de risco quando todas as covariáveis são zero. Essa função não é especificada e pode variar com o tempo, sendo a parte não paramétrica do modelo.

$\beta = (\beta_1, \beta_2, \dots, \beta_k)$ é um vetor de coeficientes de regressão que mede o efeito das covariáveis $X = (X_1, X_2, \dots, X_k)$ sobre o risco. Aqui, $\beta^T X$ é o produto escalar, ou seja, $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$.

$\exp(\beta^T X)$ é o fator multiplicativo que ajusta o risco basal com base nas covariáveis, assumindo uma relação log-linear entre as covariáveis e o logaritmo da taxa de risco.

Essa formulação implica que o efeito das covariáveis é multiplicativo no risco, e a relação entre as taxas de risco de dois indivíduos com diferentes valores de covariáveis é constante ao longo do tempo, o que é a suposição de riscos proporcionais. Por exemplo, a razão de risco (hazard ratio) entre dois indivíduos i e j é:

$$\frac{h(t | X_i)}{h(t | X_j)} = \exp(\beta^T (X_i - X_j))$$

Essa constante não depende de t , o que é a essência da proporcionalidade

A suposição fundamental do modelo de Cox é que os riscos são proporcionais, ou seja, a taxa de risco de um indivíduo é uma constante multiplicativa da taxa de risco de outro indivíduo, dependendo apenas das suas covariáveis é constante ao longo do tempo. Essa suposição é crucial para a validade do modelo e deve ser testada.

Dado que $h_0(t)$ não é especificada, não é possível escrever a verossimilhança completa do modelo como em modelos paramétricos. Cox propôs a verossimilhança parcial (partial likelihood), que permite estimar β sem precisar especificar $h_0(t)$. A verossimilhança parcial é construída considerando a probabilidade de que, dado que um evento ocorre em um certo tempo t_i , ele ocorre para o indivíduo i em vez de qualquer outro indivíduo ainda em risco. A forma é (Carvalho, M.S, et al,2011):

$$L(\beta) = \prod_{i:\delta_i=1} \frac{\exp(\beta^T X_i)}{\sum_{j \in R(t_i)} \exp(\beta^T X_j)}$$

em que:

O produto é sobre todos os indivíduos i para os quais $\delta_i = 1$, ou seja, aqueles que experimentaram o evento (observações não censuradas). Aqui, δ_i é um indicador de censura, sendo 1 se o evento ocorre e 0 se o indivíduo é censurado.

t_i é o tempo em que o indivíduo i experimentou o evento.

$R(t_i)$ é o conjunto de indivíduos em risco no tempo t_i , ou seja, todos os indivíduos que ainda não experimentaram o evento ou foram censurados antes de t_i . Isso inclui o indivíduo i e todos os outros que estão "em risco" naquele momento.

Em cada evento, escolhemos o indivíduo que experimentou o evento com base na sua contribuição relativa ao risco, dada pelas exponenciais dos produtos escalares $\beta^T X$.

Para facilitar a maximização, usa-se o logaritmo da verossimilhança parcial:

$$\ell(\beta) = \sum_{i:\delta_i=1} \left[\beta^T X_i - \log \left(\sum_{j \in R(t_i)} \exp(\beta^T X_j) \right) \right]$$

Os coeficientes β são estimados maximizando $\ell(\beta)$, o que geralmente é feito numericamente, como por métodos de Newton-Raphson, devido à complexidade do somatório no denominador.

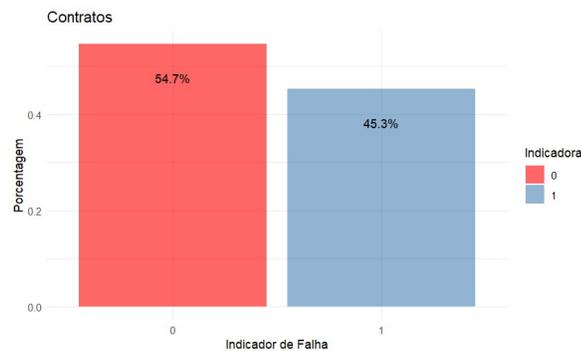
A escolha do modelo aplicável usou o teste da razão de verossimilhança e o critério de informação Akaike (AIC). O diagnóstico da adequação do modelo foi realizado pelo teste da proporcionalidade, pelos resíduos de Schoenfeld resíduos padronizados. Foi verificada a presença de observações influentes. Obtidos os coeficientes β , foram comparados os riscos entre covariáveis e calculada a razão de risco para todas as covariáveis.

6 RESULTADOS

6.1 Análise Descritiva e Exploratória

Foram consideradas falhas os registros em que se observou a interrupção do contrato durante a sua vigência, financeiramente denominado como “saída de carteira”. Considerou-se censurados todos os registros cujos contratos ainda se encontravam ativos na data de encerramento do estudo ou que tenham permanecido ativos até o seu prazo final. Os dados continham 54,7% observações censuradas, indicados por “0” enquanto 48,3% apresentaram falha, indicada por “1”.

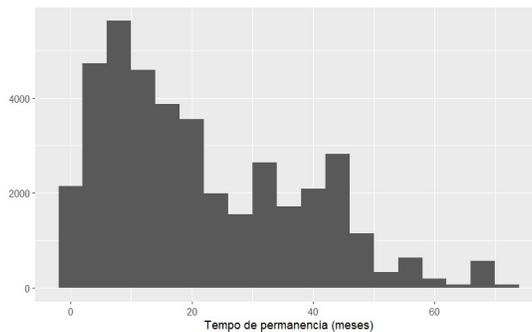
Figura 1 – Percentagem de censuras e falhas presentes no banco de dados cedidos pelo CORBAN.



Fonte: Elaborado pelo autor.

Quanto a distribuição do tempo de permanência em meses (duration) dos contratos que apresentaram falha, verifica-se uma predominância nos períodos inferiores a 24 meses, mínimo de 1 e máximo de 74 meses. Forte assimetria positiva com mediana de 18 meses e média de 20 meses, figura 2.

Figura 2 – Distribuição do nº de falhas quanto ao tempo de permanência em meses



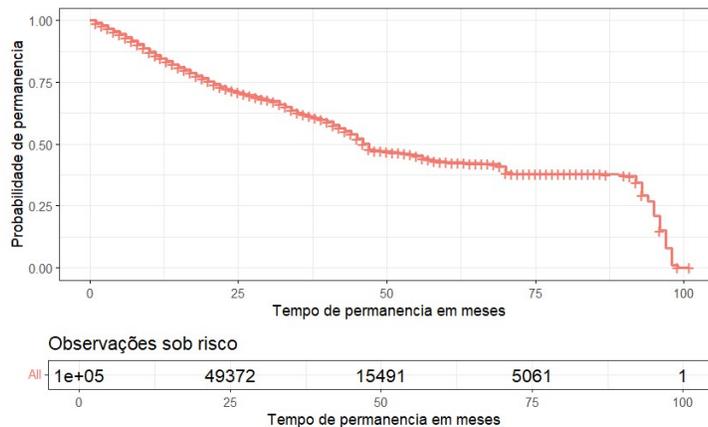
Fonte: Elaborado pelo autor.

6.2 Estimador de Kaplan- Meier

6.2.1 Geral

A estimação Kaplan-Meier sem estratificação por covariáveis pode ser visualizada na figura 3. O tempo de sobrevivência mediano (50% de sobrevivência) foi de 46 meses, aos 25 meses cerca de 70% permaneciam ativos, o nível de 25% de permanência só foi atingido no mês 94, evidenciando uma redução na taxa de risco após os 50 meses de permanência até as imediações do nonagésimo mês.

Figura 3 – Estimativa Kaplan-Meier para o tempo de sobrevivência dos contratos



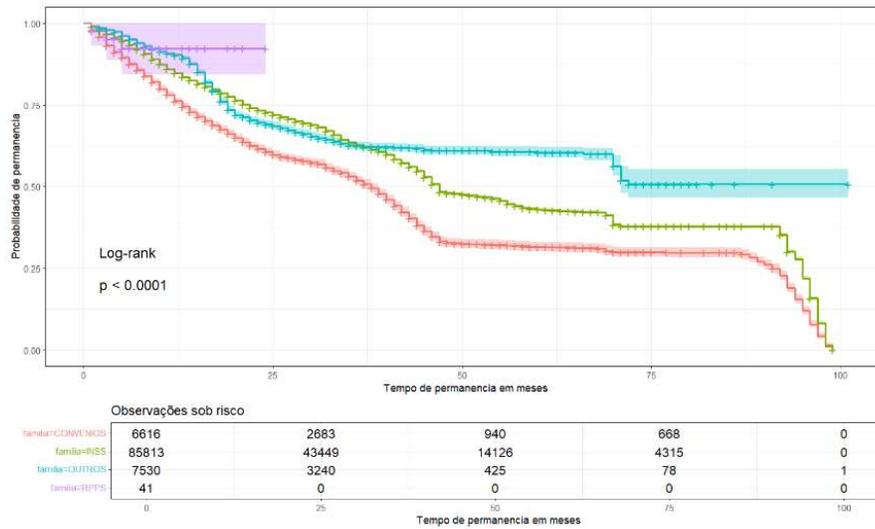
Fonte: Elaborado pelo autor.

6.2.2 Estratificada

As curvas de sobrevivência pelo EKM para cada uma das covariáveis categóricas – família, tipo proposta, linha e banco - bem como o p-valor dos respectivos teste log-rank estão nas figuras 4 a 7. O p-valor do teste log-rank refere-se a verificação da hipótese nula de que todas as categorias possuem curvas de sobrevivência idênticas. A rejeição implica que pelo menos uma delas apresenta diferença em relação às outras.

Quanto à covariável família, a categoria “ OUTROS” assume a liderança na sobrevivência a partir dos 45 meses, enquanto a categoria “ CONVENIOS” é a que apresenta menor taxa, a categoria RPPS foi introduzida recentemente pela legislação pertinente havendo censura à direita pela duração do estudo.

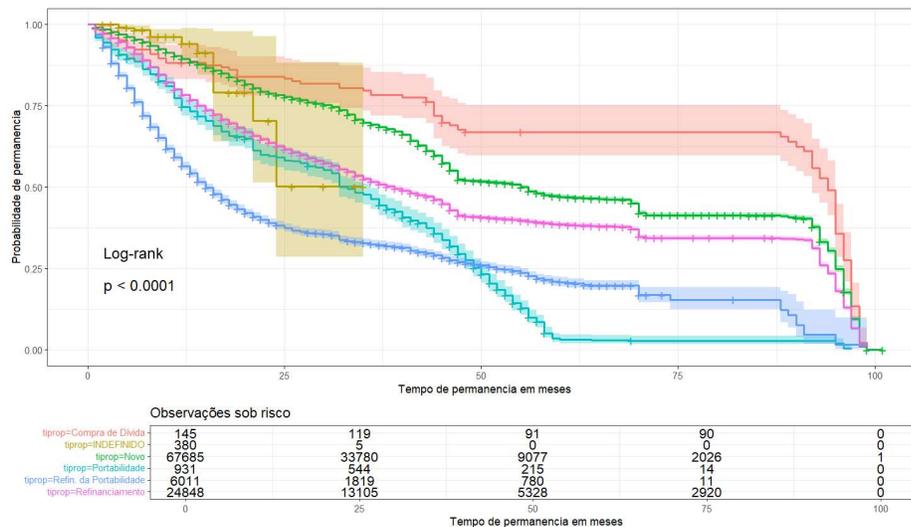
Figura 4 – EKM e log-rank estratificado para a covariável família



Fonte: Elaborado pelo autor.

Quanto à covariável tipo de proposta a categoria “Compra de dívida” tem a maior permanência. Enquanto a categoria “Refin da Portabilidade” possui maior taxa de risco com apenas 25% dos contratos sobrevivendo até os 50 meses.

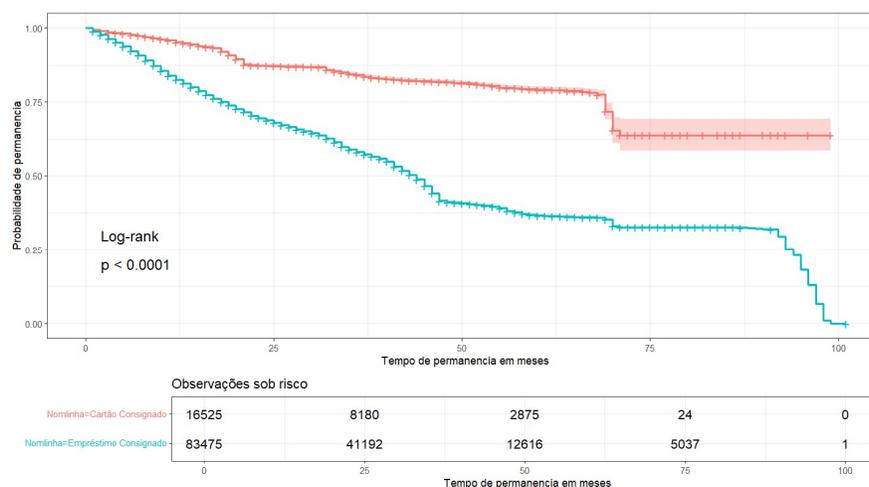
Figura 5 – EKM e log-rank estratificado para a covariável tipo proposta



Fonte: Elaborado pelo autor.

Analisando-se a linha do empréstimo o: Cartão Consignado demonstrou ser o mais durável, com mais de 50% dos contratos sem falhas até o 50º mês.

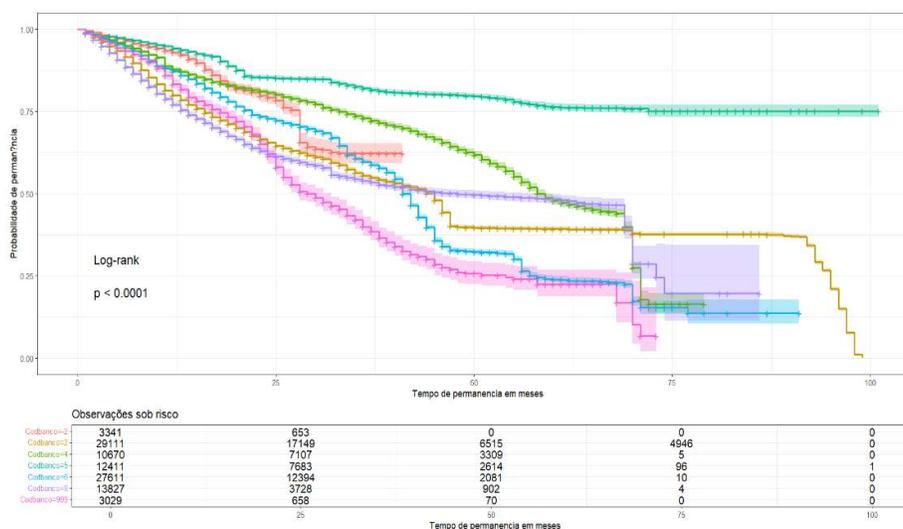
Figura 6 – EKM e log-rank estratificado para a covariável linha empréstimo



Fonte: Elaborado pelo autor.

Quanto ao Banco o de código 5 apresenta menor taxa de risco e o de código 999 a maior dentre eles.

Figura 7– EKM e log-rank estratificado para a covariável Banco



Fonte: Elaborado pelo autor.

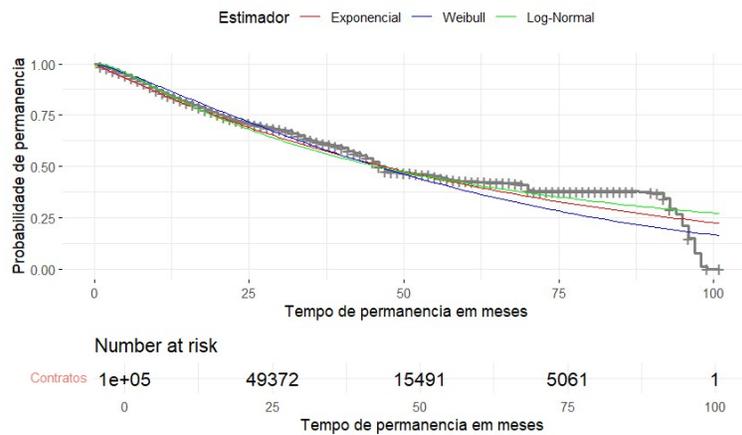
Todos os p-valores do teste log-rank foram significativos ao nível de 5%, concluiu-se pela inclusão no modelo de regressão de todas as categorias em todas as covariáveis.

6.3 Modelagem paramétrica

6.3.1 Seleção do modelo probabilístico

Calculados pelo método da máxima verossimilhança, os modelos paramétricos: exponencial, Weibull e log-normal foram plotados juntamente com o EKM na figura 8. Após os 50 meses, o modelo log normal parece melhor representar a curva de sobrevivência.

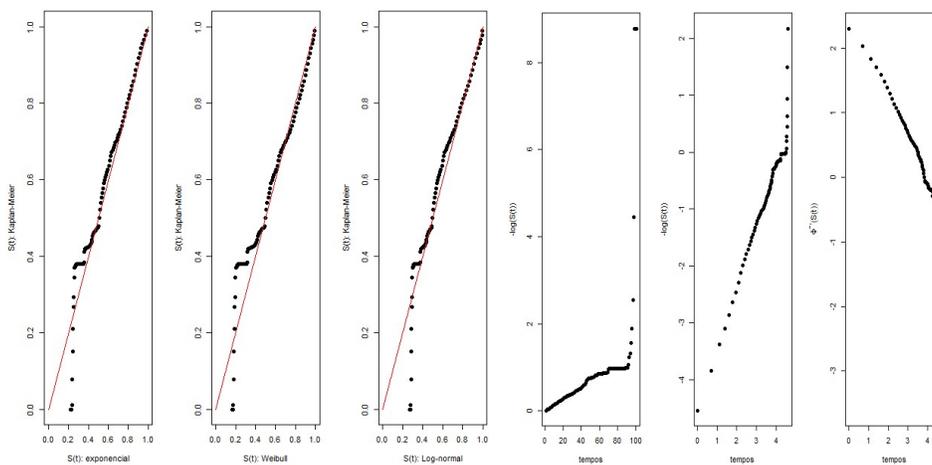
Figura 8– Modelos probabilísticos comparados com o EKM



Fonte: Elaborado pelo autor.

O teste de ajuste pelos métodos gráficos com e sem linearização das curvas indicam uma superioridade nos ajustes dos modelos Weibull e log-normal, figura 9.

Figura 9– ajuste dos modelos probabilísticos pelos métodos gráficos



Fonte: Elaborado pelo autor.

O teste de ajuste pela razão de verossimilhança comparada com a função gama generalizada rejeitou a hipótese nula com $p\text{-valor} < 0,001$ para todos os modelos, concluindo pela ausência de evidências de similaridade entre eles e a gama generalizada. Entretanto o AIC apontou o modelo Weibull como superior aos demais, valores na Tabela 6. Adotou-se o modelo Weibull como representativo da função de sobrevivência para os dados.

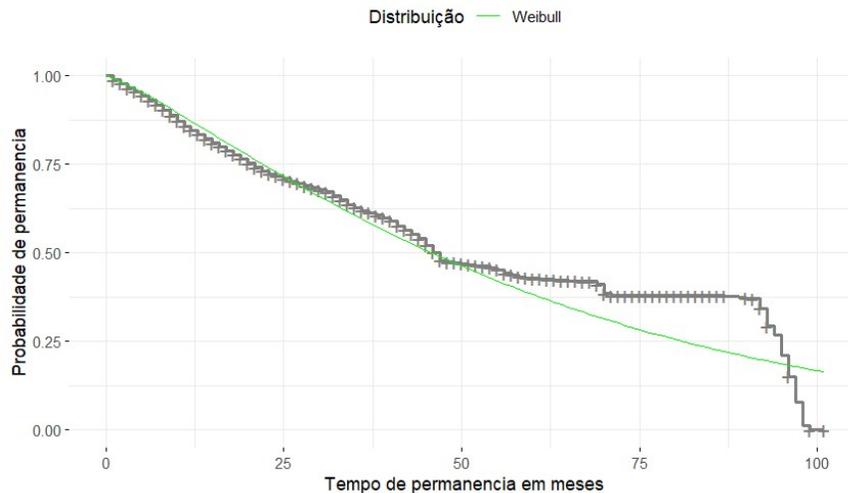
Tabela 6 - Valores do AIC para os modelos probabilísticos

Modelo	AIC
Exponencial	472.072,9
Weibull	469.726,7
Log - normal	475.038,2

Fonte: Elaborado pelo autor.

O modelo Weibull comparado ao EKM foi plotado na figura 10, ratificando o seu bom ajuste, com destaque para o período até 50 meses. Buscando-se uma modelagem para a avaliação dos ativos contratuais a valor presente, os períodos iniciais são os mais importantes.

Figura 10 – curvas de sobrevivência estimada pelos modelos Weibull e EKM



Fonte: Elaborado pelo autor.

6.3.1 Seleção do modelo de regressão

Escolhida a modelagem pela curva Weibull, foi desenvolvido de um modelo de regressão para quantificar o efeito das covariáveis do tipo

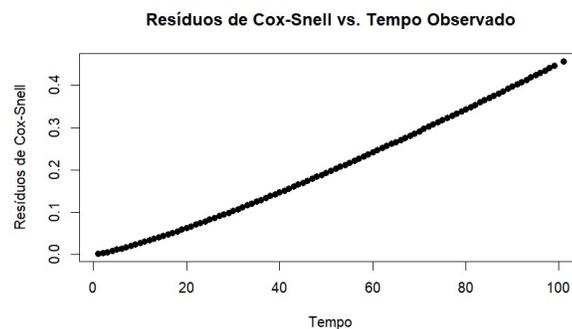
$$\text{Sobrevida da carteira} \sim \text{Tipo} + \text{Linha} + \text{Banco} + \text{Família}$$

Em busca de um modelo mais simples, foram testadas, pela razão da verossimilhança, as diversas combinações de covariáveis concluindo-se pela necessidade de manter todas a covariáveis.

6.3.2 Ajuste do modelo de regressão

A verificação da qualidade do modelo escolhido mediante os resíduos de Cox -Snell mostrou um bom ajuste com a plotagem dos valores em função do tempo assumindo uma reta de inclinação 45°, figura 11.

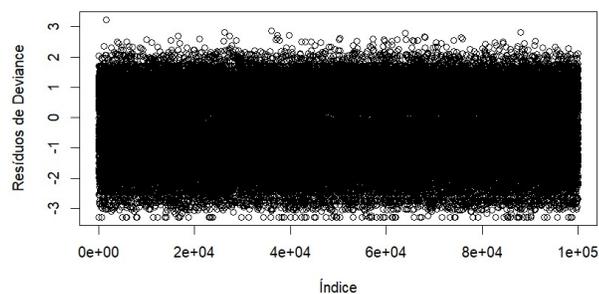
Figura 11 – Resíduos Cox-Snell versus tempo



Fonte: Elaborado pelo autor.

Os resíduos *deviance*, úteis para a detecção de outliers, mostraram uma baixa incidência de pontos afastados do “zero”, figura 12.

Figura 12 – Resíduos deviance



Fonte: Elaborado pelo autor.

O modelo Weibull ajustado tem a sua função de sobrevivência na forma

$$S(t | X) = e^{-(\lambda t)^k e^{\beta X}},$$

Assim, as categorias de covariáveis em que o $\beta > 0$ indicam um aumento da sobrevivência em relação à situação basal, inversamente se $\beta < 0$ a presença da categoria reduz a sobrevivência. Por exemplo a presença da categoria família RPPS aumenta a sobrevivência em relação a um contrato cuja família seja INSS. Tomando-se um contrato com o Codbanco6 temos a sobrevivência diminuída em relação a outro com codbanco4. Os p-valores apurados indicam que todos os coeficientes possuem significância ao nível de 5%. O scale do modelo é de 0,817; uma vez que o parâmetro de escala $\gamma = 1/\text{scale}$, então $\gamma = 1,22$. Conclui-se que a taxa de riscos é estritamente crescente em relação ao tempo. As situações basais para cada uma das covariáveis são: família – CONVENIOS, tipop – Compra de dívida, Nomlinha – CartãoConsignado, codbanco - -2.

Tabela 7 - Estimativa dos parâmetros do modelo de regressão Weibull

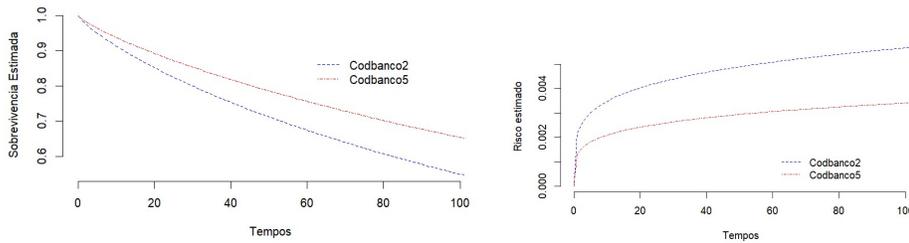
Covariável	Categoria	β	Erro -padrão	p - valor
(Intercept)		5,25528	0,08182	<2E-16
familia	INSS	0,13811	0,01379	<2E-16
familia	OUTROS	0,43925	0,02204	<2E-16
familia	RPPS	0,63864	0,47195	0,176
tipop	INDEFINIDO	0,22093	0,20598	0,2835
tipop	Novo	-0,18212	0,07041	0,0097
tipop	Portabilidade	-1,14231	0,07616	<2E-16
tipop	Refin.da Portabilidade	-0,92757	0,07129	<2E-16
tipop	Refinanciamento	-0,4405	0,07026	3,60E-10
Nomlinha	Empréstimo Consignado	-0,85764	0,02227	<2E-16
Codbanco	2	-0,16081	0,03476	3,70E-06
Codbanco	4	0,25853	0,03638	1,20E-12
Codbanco	5	0,05996	0,04035	0,1372
Codbanco	6	-0,41677	0,03467	<2E-16
Codbanco	8	-0,33928	0,03603	<2E-16
Codbanco	999	-0,5153	0,04226	<2E-16
Log(scale)		-0,20216	0,00385	<2E-16

Fonte: Elaborado pelo autor.

6.3.2 Comparando dois elementos

A modelagem permite a comparação da taxa de risco instantânea e das curvas de sobrevivência entre dois elementos cujas covariáveis sejam distintas. Exemplificando, dois contratos com a covariável banco “2” e “5” respectivamente.

Figura 13 – Curva de sobrevivência e Taxa de risco para Bancos 2 e 5



Fonte: Elaborado pelo autor.

6.4 Modelagem semiparamétrica

6.4.1 Seleção do modelo

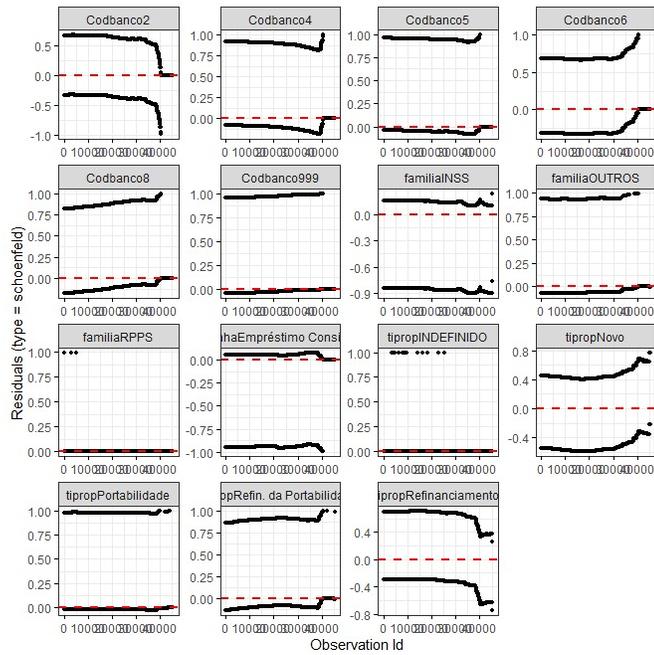
Desenvolvidos os modelos de Cox com a utilização de 4 e 3 covariáveis, aplicou-se o teste da razão de verossimilhança para comparar o modelo maior com outros menores. Todos concluíram pela rejeição da hipótese nula de equivalência entre os modelos, selecionada modelagem com 4 covariáveis:

$$\text{Sobrevida da carteira} \sim \text{Tipo} + \text{Linha} + \text{Banco} + \text{Família}$$

6.4.2 Ajuste do modelo de regressão

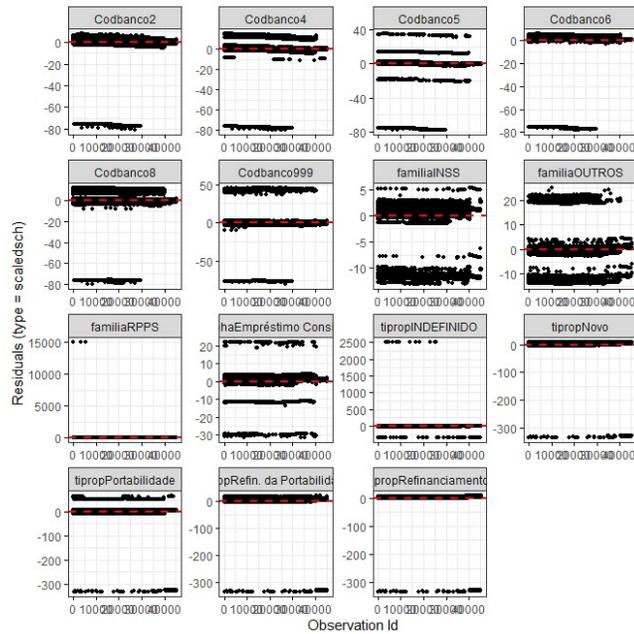
O teste da proporcionalidade dos resíduos de Schoenfeld com e sem padronização retornaram graficamente curvas aproximadamente paralelas para todas as covariáveis confirmando a hipótese de proporcionalidade, figuras 14 e 15.

Figura 14 – Coeficientes de correlação de Pearson entre os resíduos de Schoenfeld



Fonte: Elaborado pelo autor.

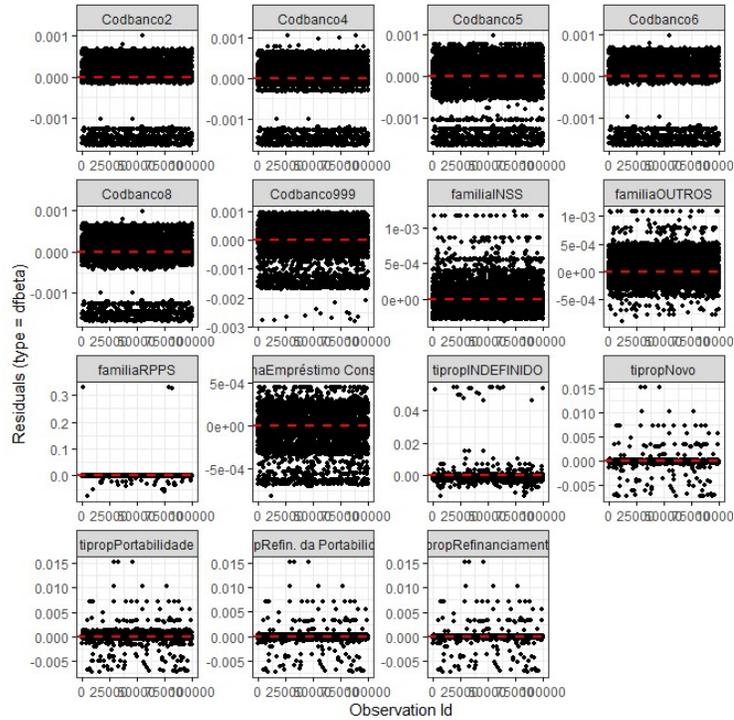
Figura 15 – Coeficientes de correlação de Pearson entre os resíduos de Schoenfeld padronizados



Fonte: Elaborado pelo autor.

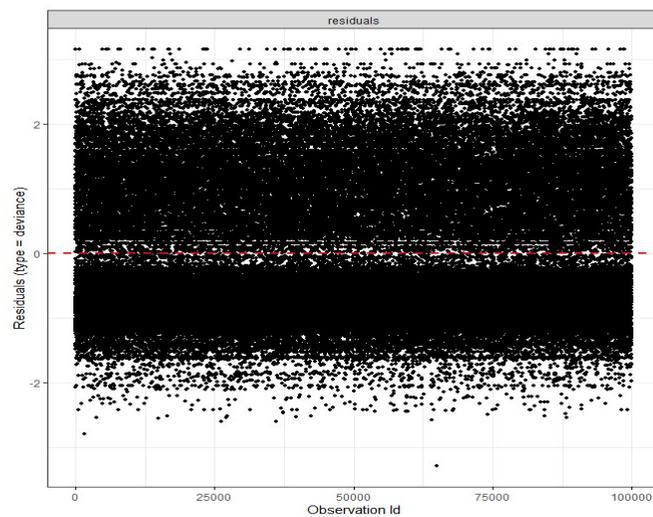
A observação de pontos aberrantes, outliers e observações influentes por meio dos resíduos escore (dfbeta) e deviance revelaram poucos pontos e sem importância para a análise, com essas características, figuras 16 e 17.

Figura 16 – Resíduos escore-dfbeta



Fonte: Elaborado pelo autor.

Figura 17 – Resíduos deviance



Fonte: Elaborado pelo autor.

Os coeficientes β da função de risco $h(t | X) = h_0(t)\exp(\beta^T X)$ do modelo estão na Tabela 8. A concordância em relação aos dados foi de 0,65. As situações basais para cada uma das covariáveis são: família – CONVENIOS, tiprop – Compra de dívida, Nomlinha – CartãoConsignado, Codbanco - -2.

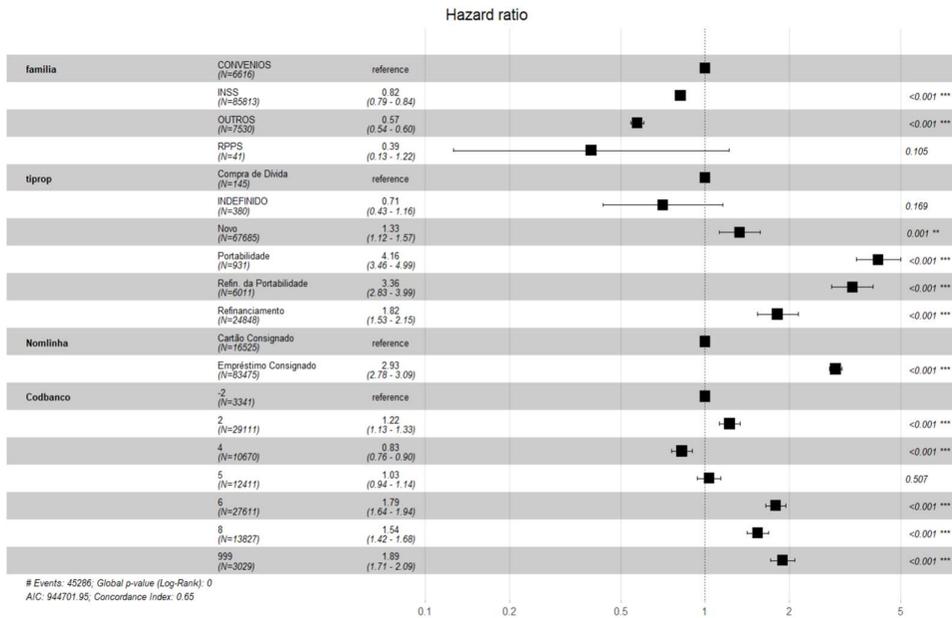
Tabela 8 – Estimativas dos coeficientes do modelo Cox

Covariavel	Categoria	β	$\exp(\beta)$	$se(\beta)$	$Pr(> z)$
familia	INSS	-0,20147	0,81753	0,01686	<2E-16
familia	OUTROS	-0,5571	0,57287	0,02687	<2E-16
familia	RPPS	-0,93625	0,3921	0,57766	0,10507
tiprop	INDEFINIDO	-0,347	0,70681	0,25211	0,1687
tiprop	Novo	0,2833	1,3275	0,08625	0,00102**
tiprop	Portabilidade	1,42506	4,15809	0,09328	<2E-16
tiprop	Refin.da Portabilidade	1,21188	3,35981	0,08742	<2E-16
tiprop	Refinanciamento	0,59619	1,81519	0,08605	4,26E-12
Nomlinha	Empréstimo Consignado	1,07468	2,92907	0,02713	<2E-16
Codbanco	2	0,20228	1,22419	0,04261	2,06E-06
Codbanco	4	-0,19097	0,82616	0,04469	1,93E-05
Codbanco	5	0,03286	1,0334	0,04956	0,50737
Codbanco	6	0,58002	1,78608	0,04232	<2E-16
Codbanco	8	0,43374	1,54302	0,04404	<2E-16
Codbanco	999	0,63785	1,89241	0,05165	<2E-16

Fonte: Elaborado pelo autor.

Excetuando-se as covariáveis: família RPPS, tiprop INDEFINIDO e codbanco5 todos os coeficientes foram significativos ao nível de 5%. O modelo aponta que a os contratos família INSS apresentam um risco de saída 0,82 vezes menor que as família CONVENIOS mantidas todas as outras covariáveis fixas. O quadro comparativo dos riscos para todas as covariáveis está na figura 18.

Figura 18 – Razão de riscos para as covariáveis

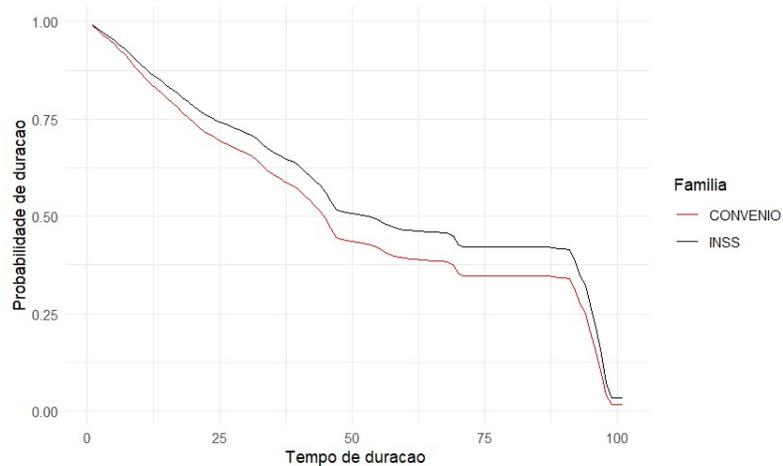


Fonte: Elaborado pelo autor.

6.3.2 Comparando dois elementos

As curvas de sobrevivência de dois contratos do banco 2, tipo = Novo, Linha= Empréstimo Consignado, no qual a família de um é INNS e do outro é CONVENIOS demonstra que o tempo mediano de duração do primeiro é maior que do segundo.

Figura 19 – Curvas de sobrevivência para família INSS e CONVENIO



Fonte: Elaborado pelo autor.

7 CONCLUSÃO

A avaliação de ativos contratuais sob risco ganha importância à medida que surgem novos instrumentos e formatações financeiras dos negócios. No caso estudado, os CORBANs, essa classe compõe a maior parte dos seus ativos necessitando de monitoramento permanente para aferir a saúde econômico-financeira da entidade empresarial. A projeção de fluxos de caixa futuros é essencial para a precificação dos produtos e garantir a liquidez ao longo do tempo.

A análise de sobrevivência dos contratos mostrou-se uma ferramenta eficaz para essa finalidade. A modelagem paramétrica pela função Weibull foi consistente com a modelagem semiparamétrica de Cox, apontando de maneira similar a direção das taxas de risco para as covariáveis. Embora os coeficientes calculados estejam com sinais opostos, trata-se, tão somente, de um efeito do algoritmo computacional utilizado, mas cuja interpretação leva às mesmas conclusões.

Desenvolvimentos ulteriores poderão ser feitos, tais como:

- a) cálculo do valor presente dos fluxos de caixa esperado de um determinado contrato, sabendo-se que ele se encontra ativo numa idade qualquer;
- b) cálculo da margem bruta de lucro esperada para um contrato em sua data inicial;
- c) cálculo dos valores de comissão a ser paga para a obtenção do contrato.

Algumas técnicas de análise também merecem ser avaliadas para contornar situações práticas de mudança das covariáveis ou da estrutura dos dados com o tempo, suavização das curvas, eventos múltiplos ou competitivos e modelos de fragilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, B. H. **Análise de sobrevivência aplicada a contratos de serviços: uma abordagem estatística para identificação de riscos**. Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, 2019.
- ALLISON, Paul D. *Survival analysis using SAS: a practical guide*. 2. ed. Cary: SAS Institute, 2010.
- CARVALHO, Marília Sá et al. *Análise de sobrevivência: teoria e aplicações em saúde*. 2. ed. Rio de Janeiro: Fiocruz, 2011.
- COCK, R. J.; LAWLESS, J. F. **The statistical analysis of recurrent events**. New York: Springer, 2007.
- COLOSIMO, Enrico Antônio; GIOLO, Suely Ruiz. *Análise de sobrevivência aplicada*. São Paulo: Blucher, 2006.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. 2. ed. São Paulo: Editora Blucher, 2017.
- COX, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 34, n. 2, p. 187–220, 1972.
- COX, D. R.; OAKES, D. **Analysis of survival data**. London: Chapman & Hall, 2004.
- COX, David R.; OAKES, David. *Analysis of survival data*. London: Chapman & Hall/CRC, 2001.
- DELOITTE. *IFRS 15: revenue from contracts with customers - a closer look*. London: Deloitte Touche Tohmatsu Limited, 2020.
- DICKSON, David C. M.; HARDY, Mary R.; WATERS, Howard R. **Actuarial mathematics for life contingent risks**. 2. ed. Cambridge: Cambridge University Press, 2013.
- DUDLEY, W. N.; WICKHAM, R.; COOMBS, N. An introduction to survival statistics: Kaplan-Meier analysis. *Journal of Advanced Practice in Oncology*, v. 7, n. 1, p. 91-100, Jan./Feb. 2016. DOI: 10.6004/jadpro.2016.7.1.8.
- EY. *Applying IFRS 15: a practical guide to revenue recognition*. New York: Ernst & Young Global Limited, 2021.
- FLEMING, T. R.; HARRINGTON, D. P. **Counting processes and survival analysis**. 2. ed. New York: Wiley, 2001.
- GIOLO, Suely Ruiz. *Análise de sobrevivência aplicada*. 2. ed. São Paulo: Blucher, 2014.

HOSMER, David W.; LEMESHOW, Stanley; MAY, Susanne. *Applied survival analysis: regression modeling of time-to-event data*. 2. ed. Hoboken: Wiley, 2008.

KALBFLEISCH, John D.; PRENTICE, Ross L. **The statistical analysis of failure time data**. 2. ed. Hoboken: Wiley, 2002.

KASSAMBARA, A., Kosinski, M., Biecek, P., & Fabian, S. (2024). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.5.0. CRAN

KPMG. *Revenue from contracts with customers: IFRS 15 handbook*. Amstelveen: KPMG International, 2018.

LAWLESS, Jerald F. **Statistical models and methods for lifetime data**. 2. ed. Hoboken: Wiley, 2003.

LEE, Elisa T.; WANG, John Wenyu. *Statistical methods for survival data analysis*. 3. ed. Hoboken: Wiley, 2003.

MACHIN, D.; CHEUNG, Y. B.; PARMAR, M. K. B. *Survival analysis: a practical approach*. 2. ed. Chichester: Wiley, 2006.

NIIYAMA, Clóvis Augusto. Análise clássica e bayesiana do modelo Weibull modificado generalizado. 2013. 130 f. Dissertação (mestrado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Ciências e Tecnologia, 2013.

OLIVEIRA, Márcio A.; LOUZADA-NETO, Francisco. **Modelos de sobrevivência aplicados a dados de seguro**. *Revista Brasileira de Estatística*, v. 65, n. 1, p. 45-60, 2004.

PINHEIRO, Hildete P.; NASCIMENTO, Diego C. **Bayesian survival analysis: recent advances**. *Brazilian Journal of Probability and Statistics*, v. 30, n. 4, p. 599-620, 2016.

PwC. *IFRS 15: revenue recognition - practical insights*. London: PricewaterhouseCoopers, 2019.

SANTOS, L. T. M. **Modelagem de risco em contratos de serviços utilizando análise de sobrevivência**. Dissertação de Mestrado. Universidade Estadual Paulista, 2022.

SILVA, M. G. **Análise de sobrevivência e técnicas atuariais aplicadas à gestão de contratos de serviços**. Tese de Doutorado. Universidade de São Paulo, 2021.

SOUZA, João Paulo; LIMA, Maria Fernanda. **Aplicação do IFRS 15 no Brasil: desafios e perspectivas**. *Revista Contabilidade & Finanças*, v. 33, n. 90, p. 123-138, 2022.

SOUZA, J.M.P. - **Análise de Sobrevivência: Conceitos e Aplicações**. Editora UFPR, 2009.

THERNEAU, T. M. A Package for Survival Analysis in R [recurso eletrônico]. R package version 3.7-0, 2024. Disponível em: <https://CRAN.R-project.org/package=survival>. Acessado em: [data de acesso] (CRAN).

THERNEAU, T. M.; GRAMBSCH, P. M. Modeling survival data: extending the Cox model. New York: Springer, 2000.