



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Modelo de previsão de risco de crédito com atribuição do limite através do lucro previsto

Autora: Andressa Bruna Costa
Orientadora: Professora Dra. Lisiane Priscila Roldão Selau

Porto Alegre, 03 de julho de 2015.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Modelo de previsão de risco de crédito com atribuição do limite através do lucro previsto

Autora: Andressa Bruna Costa

Trabalho de Conclusão de Curso
apresentado para obtenção do
grau de Bacharel em Estatística.

Banca Examinadora:
Prof^a Dr^a Lisiane Selau
Bel. Sabrina Zanata Grebin

Porto Alegre, 03 de julho de 2015.

"Existem dois tipos de riscos: Aqueles que não podemos nos dar ao luxo de correr, e aqueles que não podemos nos dar ao luxo de não correr".

Peter Drucker

Agradecimentos

Aos meus pais, a minha infinita gratidão, pelo amor mais puro e sincero, por todo apoio, incentivo e dedicação para a estabilização emocional e viabilização financeira dos meus passos.

Ao Andi e ao Akauã, por serem mais que irmãos, amigos que tenho o orgulho de compartilhar o sangue.

Ao Alexandre, pelo carinho, pela intimidade compartilhada e pela compreensão nos momentos de ausência.

À Nati, irmã de coração que desempenhou tantos papéis importantes nessa jornada.

Às amigas queridas que mesmo depois de tanto tempo, se fazem presentes mesmo distantes: Sabri, Bruna, Carla, Gabi, Anna e Thaís.

Às amigas mais recentes, mas que já compartilharam comigo muitas emoções e aventuras: Ana, Pâmela e Daia.

Aos colegas e ex-colegas que fizeram parte da minha formação, principalmente aos que, despretensiosamente, contribuíram para que fosse possível chegar até aqui.

À profe Lisi, que me inspirou e aceitou meu pedido para orientação deste trabalho e me atendeu prontamente em todas as dificuldades.

A todos que, de uma forma ou outra, passaram pelo meu caminho e, especialmente, aos que permaneceram nele, o meu muito obrigada.

Resumo

O objetivo do presente trabalho é propor um modelo de previsão de crédito, através de regressão linear múltipla, em que a decisão da concessão é uma baseada na medida monetária do lucro esperado por cada proponente. Neste modelo, são reprovados aqueles com os quais se espera uma medida de lucro menor que zero, ou seja, prejuízo e, por outro lado, por representarem ganho, sugerir a atribuição do limite de crédito somente aos aprovados pelo modelo, de forma condizente com a medida monetária de lucro esperado. O desenvolvimento do modelo consiste de três grandes etapas: 1) pré-processamento, 2) construção e avaliação do modelo, e 3) sistemática para a atribuição do limite. Na primeira etapa, são desenvolvidos os passos: (i) delimitação da população, (ii) seleção da amostra, e (iii) análise preliminar. A segunda etapa envolve os passos: (i) construção do modelo e (ii) avaliação da qualidade do modelo. A terceira e última etapa consiste da avaliação do lucro esperado e da atribuição do limite equivalente ao valor de venda necessário para se observar tal lucro. O estudo envolveu dados reais de concessão de crédito de uma rede de farmácias. De forma a identificar o aumento potencial nos ganhos através da utilização do modelo de previsão, são avaliados os cenários anterior e posterior à implementação do modelo, que demonstra uma inversão de resultados, passando de um prejuízo a lucro, mesmo com menos da metade dos clientes aprovados em relação ao cenário de aprovação de todos os clientes. Também é realizada uma análise de sensibilidade sugerindo que mesmo aprovando aqueles com os quais se espera prejuízo, acima de determinada faixa, o resultado acumulado continua sendo positivo. Nesse sentido, o modelo de previsão com variável resposta contínua, que avalia o lucro esperado, mostra-se uma ferramenta efetiva na concessão de crédito e atribuição do limite.

Palavras-chave: Modelo de previsão – Regressão linear múltipla – Decisão monetária – Atribuição de limite

SUMÁRIO

1	INTRODUÇÃO	7
2	FUNDAMENTAÇÃO TEÓRICA	9
2.1	<i>CREDIT SCORING</i>	9
2.2	LIMITE DE CRÉDITO	10
2.3	REGRESSÃO LINEAR MÚLTIPLA	11
3	MODELO PROPOSTO	12
3.1	PRÉ-PROCESSAMENTO	13
3.1.1	Delimitação da população	13
3.1.2	Seleção da amostra	14
3.1.3	Análise preliminar	14
3.2	CONSTRUÇÃO E AVALIAÇÃO DO MODELO	15
3.2.1	Construção do modelo	15
3.2.2	Qualidade do modelo.....	16
3.3	SISTEMÁTICA PARA A ATRIBUIÇÃO DO LIMITE	16
4	RESULTADOS	17
4.1	PRÉ-PROCESSAMENTO	17
4.1.1	Delimitação da população	17
4.1.2	Seleção da amostra	18
4.1.3	Análise preliminar	20
4.2	CONSTRUÇÃO E AVALIAÇÃO DO MODELO	22
4.2.1	Construção do modelo	22
4.2.2	Qualidade do modelo.....	25
4.3	ATRIBUIÇÃO DO LIMITE.....	30
5	CONSIDERAÇÕES FINAIS	33
6	REFERÊNCIAS BIBLIOGRÁFICAS	35

1 INTRODUÇÃO

O crescimento da oferta de crédito à pessoa física nos últimos tempos tem impulsionado o comércio de produtos e serviços. As empresas têm investido nessa oportunidade através de crediários próprios que, além de fidelizar o cliente, também impulsionam as vendas do varejo e a absorção das taxas de juros de parcelamento. Esse crédito também tem estado exposto cada vez mais às classes mais baixas.

Em 2014, a DIEESE publicou uma nota técnica sobre a evolução do crédito na economia brasileira de 2008 a 2013, em que afirma que a ampliação do mercado de crédito é uma das razões que explicam o crescimento da economia brasileira no período recente. Em dezembro de 2002, a relação crédito/PIB era de 23,8%, passando a 55,8% em fevereiro de 2014. A evolução do crédito total e à pessoa física em relação ao PIB nos últimos anos é apresentado na Figura 1.

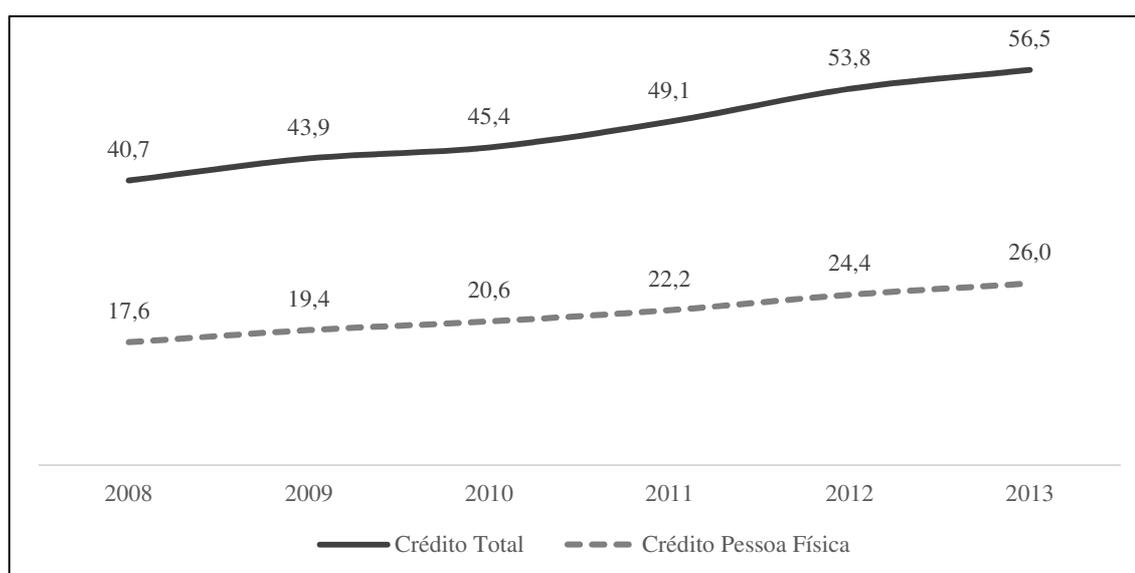


Figura 1. Evolução do crédito em relação ao PIB (%)

Fonte: DIEESE

Para que as empresas possam expor esse crédito, é também necessário que se avalie todas as propostas igualmente e, nesse sentido, torna-se essencial a utilização de ferramentas rápidas e eficazes que auxiliem na tomada de decisão. Neste cenário, os modelos de *credit scoring*, que classificam os proponentes a crédito quanto ao seu risco como cliente, estão cada vez mais importantes e necessários.

Apesar disso, os sistemas de crédito amplamente praticados atualmente avaliam apenas a probabilidade do proponente a crédito se tornar inadimplente até determinado momento, sem mensurar se, mesmo neste cenário, o lucro obtido com ele compensaria a perda. Assim, se estiver aquém do ponto de corte praticado pela empresa, seu crédito será rejeitado, não importando quantos pagamentos e quanto lucro se poderia ter com ele antes do sinistro. Desta forma, surgem novas incertezas por parte do mercado, que está cada vez mais dinâmico e competitivo.

Os modelos de concessão de crédito podem ir muito além de encaixar, através de sua pontuação resultante, o proponente a crédito no grupo de bons pagadores, concedendo-lhe crédito, ou no de maus pagadores, recusando-o. Gonçalves (2005) afirma que, na prática, as instituições utilizam este conceito devido à maior facilidade de trabalhar com modelos de resposta binária. Segundo Selau (2012), a utilização de uma escala dicotômica na definição do desempenho dos clientes quanto à inadimplência constitui perda de informação, e sugere que escalas contínuas para determinar o comportamento de pagamento possam ter melhor aproveitamento de informação.

Thomas (2000) argumenta que, ao invés de procurar minimizar o percentual de clientes que não pagará, as empresas estão esperando poder identificar os clientes que são mais lucrativos. Acrescenta, ainda, que há um problema em utilizar informações de clientes anteriores para prever a performance de um novo proponente: quando negado, a companhia não terá conhecimento sobre como teria sido seu comportamento. Dessa forma, se a empresa afirma que os rejeitados pelo modelo seriam maus pagadores, esta decisão se perpetuará em qualquer sistema de pontuação baseado nestes dados, e alguns grupos de potenciais clientes nunca terão a oportunidade de provar que merecem crédito.

Assim, pode-se resumir o problema de pesquisa em duas questões:

- i. Como identificar os clientes que podem gerar lucro para a empresa e conceder-lhes crédito?
- ii. Como determinar o limite de crédito adequado a ser concedido através do lucro previsto?

Neste sentido, o objetivo deste trabalho é propor um modelo de concessão de crédito cuja variável resposta, que indicará a concessão ou não de crédito, será uma medida monetária contínua, em que valores maiores que zero representam lucro e, menores que zero, prejuízo. Além disso, pretende-se determinar, adequadamente, aos clientes com lucro esperado maior que o zero, o seu limite de crédito.

O presente trabalho está dividido em cinco seções, incluindo esta primeira, onde são apresentadas as ideias iniciais e objetivos do estudo. Na próxima seção, é apresentada a fundamentação teórica dos conceitos importantes para o entendimento das técnicas utilizadas e resultados obtidos. A terceira seção trata da sistemática para o desenvolvimento do modelo proposto e para a atribuição do limite e a quarta, dos resultados obtidos para o modelo proposto e para a determinação do limite de crédito. Por fim, a quinta e última seção apresenta as conclusões obtidas com o estudo.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CREDIT SCORING

Credit Scoring é definido por Thomas (2000) como um meio para reconhecer diferentes grupos numa população quando se consegue enxergar as características que relacionam os grupos. Com a chegada dos cartões de crédito no final dos anos 60 e seu consecutivo aumento nas solicitações de crédito aos bancos e outros emissores, a decisão da concessão ou recusa, que até o momento era realizada subjetivamente por analistas, proposta a proposta, passou a ser vista como ultrapassada e, então, foi substituída por métodos automáticos. Não demorou para se perceber que o novo sistema tinha melhor desempenho, em termos de tempo e predição, do que qualquer julgamento humano.

Modelos de análise para concessão de crédito, conhecidos como *credit scoring*, baseiam-se em dados históricos da base de clientes existentes para avaliar se um futuro cliente terá mais chances de ser um bom ou mau pagador, medindo a propensão de tornar-se inadimplente (GONÇALVES, 2005). Através desses dados, métodos estatísticos geram uma pontuação que represente uma medida de risco associada ao tomador (SELAU, 2012). A esta pontuação, dá-se o nome de *score*.

Esta técnica busca maximizar a concessão de crédito aos clientes que honrarão suas dívidas, aceitando mais bons pagadores, e minimizar a concessão àqueles que causarão inadimplência, rejeitando mais maus pagadores. Dessa forma, o objetivo torna-se o equilíbrio entre mais vendas e menos risco.

Além dos modelos de *credit scoring* tradicionais, que utilizam a regressão logística para mensurar a propensão do cliente se tornar inadimplente, é crescente o uso de diferentes ferramentas estatísticas, como análise discriminante, árvores de decisão e redes neurais. Selau

(2012) aprofunda os estudos em modelos combinados (*ensemble*), que utilizam diversas técnicas agregadas para chegar a um resultado melhor que os obtidos individualmente, e também em modelos em dois estágios (*hybrid*), que utilizam a resposta de um primeiro método para compor um segundo modelo de natureza diferente para a mesma população.

Steiner *et al.* (1999) destacam que qualquer erro na decisão de conceder o crédito pode significar que, em uma única operação, haja a perda do ganho obtido em dezenas de outras transações bem-sucedidas, já que o não recebimento representa a perda total do montante emprestado.

Os mesmos sistemas estão sendo usados além da concessão de crédito. Empresas utilizam-os para identificar os clientes conforme seu perfil de compra, e determinar quais são os mais propensos a serem impactados com lançamentos de novos produtos, quais continuariam utilizando o produto após o período de lançamento e até mesmo quais têm maior propensão a buscar crédito na concorrência (THOMAS, 2000). Assim, as empresas podem direcionar suas campanhas de forma a incentivar o consumo dos produtos certos aos clientes certos.

2.2 LIMITE DE CRÉDITO

O limite de crédito à pessoa física é a definição do valor máximo que o concessor admite entregar ao cliente, em forma de produtos, serviços ou do próprio valor em espécie, diante da avaliação das suas características e do seu potencial de devolução de todo valor tomado dentro do prazo estipulado (SILVA, 2002).

Bancos e outras instituições de crédito possuem maneiras distintas para determinar o valor do limite de cada cliente. Sá (2004) sugere que, para isso, sejam utilizadas uma das alternativas a seguir: i) limite de crédito múltiplo da renda comprovada do cliente, ii) limite de crédito baseado na média histórica de compra do cliente, ou iii) limite de crédito calculado com base em percentual da renda do cliente e no prazo máximo de financiamento da empresa, em que a empresa deve determinar o valor máximo das prestações que o cliente pode assumir, com base no percentual máximo de comprometimento de sua renda com o pagamento das prestações e o número máximo de prestações do plano de financiamento, obtendo o valor final multiplicando o valor máximo da prestação pelo número de prestações possíveis.

Atualmente, uma opção tem sido difundida entre os concessores, que é o cálculo da renda presumida do cliente através de suas características, cruzada com sua faixa de score determinado pelo modelo de crédito. Esta alternativa considera o menor valor entre a renda

cadastrada pelo cliente (comprovada ou não) e a renda presumida calculada. O limite final é determinado pelo percentual sobre a menor renda, que varia conforme a faixa de score, onde a maior faixa pode até mesmo estar acima do valor da renda (DINIZ; LOUZADA NETO, 2011).

Um desafio das empresas é oferecer a quantidade certa de recursos ao proponente a crédito. Independente da alternativa definida e do valor concedido, no ato da concessão todas as suposições sobre o novo cliente são baseadas em comportamentos de tomadores passados, e mesmo com propensão a ter a performance semelhante ao seu perfil, cada caso pode ter desempenho melhor ou pior do que o esperado. Portanto, todo limite deve ter um tempo de maturação e ser recalculado posteriormente, desta vez com informação referente ao seu comportamento, e não mais aos dos semelhantes. Apesar disso, Silva (2002) considera essa decisão pessoal, não podendo ser tomada com base exclusivamente analítica.

2.3 REGRESSÃO LINEAR MÚLTIPLA

A análise de regressão linear múltipla, apesar de não ser a mais utilizada na construção de modelos de crédito, é um dos métodos mais utilizados entre todos os métodos estatísticos (NETER *et al.*, 2005). Seu objetivo é encontrar relações lineares entre uma variável dependente (resposta) e variáveis regressoras. Estas relações são obtidas através da Equação 1.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon \quad (1)$$

onde y representa a variável resposta e os parâmetros β_j , $j = 0, 1, \dots, k$, são os coeficientes de regressão, que representam o incremento esperado na resposta y por unidade de x_j , quando todas as demais variáveis regressoras x_j ($i \neq j$) são mantidas constantes, correspondendo a acréscimo no valor da variável resposta, quando o sinal de β_j é positivo, ou decréscimo, quando o sinal é negativo. O termo β_0 é o intercepto que representa o valor de y quando todas as parcelas β_kx_k são nulas e ε representa o termo de erro aleatório de y (MONTGOMERY *et al.*, 2006). O método de mínimos quadrados ordinários (MQO), proposto por Gauss e Legendre, que consiste em minimizar a soma de quadrados das diferenças entre as observações y_i e suas estimativas, é o principal método utilizado para a estimação dos parâmetros β_j .

A aplicação da regressão linear múltipla está sustentada no atendimento de quatro suposições do modelo. São elas: i) variância constante dos termos de erro

(homoscedasticidade); ii) independência dos erros; iii) normalidade dos erros; e iv) ausência de multicolinearidade entre as variáveis preditoras.

Para verificar a suposição de homoscedasticidade dos termos de erro, pode-se visualizar graficamente a dispersão entre os resíduos e os valores preditos, onde procura-se observar uma distribuição aleatória em torno de zero, sem tendências ou formato aparente. Por meio desta verificação, também se observa a propriedade de que a média dos resíduos é igual a zero. Além disso, é possível verificar estas suposições através do teste de Breusch-Pagan ou do teste de Cook-Weisberg (GREENE, 2003).

A suposição de independência dos erros pode ser observada pelo teste de Durbin-Watson (GUJARATI E PORTER, 2009). Para isso, encontra-se a estatística de teste d_U e conclui-se que os erros são independentes para valores que sejam maiores que o valor tabelado d_U , porém iguais ou menores que $(4 - d_U)$, onde d_U é valor crítico tabelado do teste, variando conforme a significância e tamanho da amostra. Para verificar a suposição de que os erros possuem distribuição normal, o diagnóstico mais simples é um histograma dos resíduos, mas também pode ser analisada através do gráfico de probabilidade normal ou do teste de Shapiro-Wilk, proposto por Shapiro & Wilk (1965).

A última verificação necessária é a de ausência de multicolinearidade, que acontece quando duas ou mais variáveis explicativas são altamente correlacionadas, e sua existência no modelo tende a produzir estimativas de coeficientes de regressão inflacionadas em valor absoluto (MONTGOMERY *et al.*, 2006). Uma maneira de detectar a existência de multicolinearidade entre as variáveis preditoras é através do cálculo do fator de inflação da variância (VIF), que mensura o efeito da dependência entre regressores na variância de cada termo. Montgomery *et al.* (2006) sugerem que apenas valores resultantes de VIF maiores que 10 para uma ou mais variáveis explicativas implicam em coeficientes de regressão mal estimados por causa da multicolinearidade.

3 MODELO PROPOSTO

Nesta seção, será apresentada a sistemática para a construção do modelo proposto, composta pelas etapas de pré-processamento (delimitação da população, seleção da amostra e análise preliminar), desenvolvimento do modelo de previsão do lucro (construção do modelo e verificação da qualidade do modelo) e atribuição de limite.

3.1 PRÉ-PROCESSAMENTO

A etapa de pré-processamento consiste da coleta e do tratamento dos dados, contemplando três passos: a delimitação da população, a seleção da amostra e da análise preliminar das observações.

3.1.1 Delimitação da população

A construção de um modelo de previsão de crédito baseia-se na avaliação do resultado de créditos previamente concedidos por uma empresa. Para que se possa avaliar os resultados e desenvolver o modelo para este negócio, os conceitos de desempenhos aceitáveis devem ser definidos. O primeiro passo é o cálculo do Lucro Observado, que considera a margem de lucro sobre as vendas e os valores pagos pelos clientes, obtido a partir da Equação 2.

$$\text{Lucro Observado} = \text{Margem de Lucro} * \text{Valor Pago} - \text{Valor Não Pago} \quad (2)$$

onde Margem de Lucro é uma medida percentual obtida pela relação entre todo o valor recebido pela empresa por meio de suas vendas e todo valor gasto para a compra dos produtos e manutenção do negócio, Valor Pago representa o valor total pago pelo cliente com pouco ou nenhum atraso e Valor Não Pago é o valor total em aberto ou pago com atraso considerável, assumindo que pagamentos após este tempo não são contabilizados como lucrativos, apenas pagam os gastos com cobrança e negativação decorrentes da inadimplência. O concessor é quem deve definir os conceitos de pouco atraso e de atraso considerável, baseado no que está disposto a aceitar no seu negócio.

Após o cálculo do lucro observado, todos os clientes que tiveram crédito concedido são separados em quatro grupos: i) sem uso – clientes sem utilização do crédito; ii) bons – clientes com lucro esperado maior que zero; iii) maus – clientes com lucro esperado menor que zero e iv) indefinidos – clientes que utilizaram o crédito, porém possuem apenas parcelas a vencer.

Segundo Diniz e Louzada Neto (2012), é fundamental que a forma como os dados se relacionaram com o desempenho do crédito no passado seja similar no futuro. Além disso, toda informação utilizada na criação do modelo, deve ser oriunda dos registros cadastrais, transacionais e de pagamento fornecidas pela empresa e, portanto, os resultados obtidos a partir dela devem ser utilizados exclusivamente para este negócio.

3.1.2 Seleção da amostra

Para que se possa ter, além de dados cadastrais em número suficientemente grande para a modelagem, também a verificação dos clientes perante seu desempenho de pagamento, Lawrence (1992) sugere que os mesmos sejam consolidados num período de 12 a 18 meses após a concessão, para que, assim, observe-se a ocorrência de pagamentos e atrasos, contribuindo para a definição de bons e maus pagadores.

Uma forma de testar o poder preditivo do modelo antes de implementá-lo é medir a resposta dos clientes que não fizeram parte da construção do modelo para verificar se o desempenho se estende a novas observações provenientes da mesma população, separando os dados em duas amostras. Haykin (2001) sugere a partição dos dados, sendo uma parte para análise, com 80% das observações, que servirá para a construção do modelo e outra, com 20%, que servirá para testar o modelo construído.

3.1.3 Análise preliminar

Mesmo que as variáveis estejam limitadas à disponibilidade dos registros, é provável que nem todas ajudem a explicar a variável resposta ou, ainda, que algumas não possam ser utilizadas no seu formato original, porém possam ser combinadas com outras ou transformadas em novas variáveis. É imprescindível que, antes de qualquer análise, se faça a verificação da qualidade dos dados, atuando na correção de possíveis erros de preenchimento, padronização de respostas, eliminação de dados inconsistentes e verificação da existência e tratamento de dados faltantes (*missing*).

Para determinar, dentre as variáveis disponíveis, as que integrarão o modelo como preditoras, utilizam-se as tabelas de contingência para calcular os riscos relativos (RR), dividindo-se o percentual de clientes do grupo bom pelo percentual do grupo mau de todos os níveis de cada variável, conforme a Equação 3.

$$RR_k = \frac{\%Bons_k}{\%Maus_k}, \text{ onde } \%Bons_k = \frac{Bons_k}{Bons} \times 100 \text{ e } \%Mau_k = \frac{Maus_k}{Maus} \times 100. \quad (3)$$

Lewis (1992) propõe, como regra geral, o agrupamento dos riscos relativos em 7 classes, do péssimo ao excelente, conforme os intervalos apresentados na Tabela 1.

Tabela 1. Classes de risco relativo para agrupamento

Classe	Péssimo	Muito mau	Mau	Neutro	Bom	Muito bom	Excelente
Intervalo RR	<0,50	[0,50;0,67)	[0,67;0,90)	[0,90;1,10)	[1,10;1,50)	[1,50; 2,00)	$\geq 2,00$

Os atributos com RR neutro não diferenciam bons e maus clientes, pois a relação de bons para maus é muito próxima de 1 bom para 1 mau e, por isso, estes casos não serão incluídos na construção do modelo. As demais, serão agrupadas conforme sua classe de risco relativo.

Como grande parte das variáveis cadastrais são nominais ou categóricas, elas devem ser transformadas de modo a evitar possíveis problemas causados pela não linearidade dos atributos. Para as variáveis categóricas, é criada uma variável *dummy* para cada categoria, que só assumem os valores 1, quando possui a característica, ou 0, quando não possui. No caso das variáveis com muitas categorias, as *dummies* são criadas de acordo com o agrupamento do risco relativo.

3.2 CONSTRUÇÃO E AVALIAÇÃO DO MODELO

Esta etapa consiste da apresentação da sistemática para a construção e avaliação da qualidade do modelo proposto.

3.2.1 Construção do modelo

A proposta para a construção de um modelo de crédito, que estima o lucro esperado por cliente ao invés da propensão de o mesmo se tornar inadimplente, traz a necessidade de trabalhar com um modelo com resposta contínua. Neste caso, será utilizada a Regressão Linear Múltipla, em que deseja-se estimar o lucro (ou prejuízo) do proponente a crédito em função das variáveis cadastrais disponíveis (preditoras).

O método de seleção de variáveis utilizado é o *stepwise*, um método automático, que inicia o modelo com a variável que possui maior correlação com a variável resposta e, iterativamente, inclui e exclui as variáveis conforme seu desempenho, chegando, por fim, a um modelo com apenas as variáveis que são significativas e influenciam o resultado da variável dependente.

O bom desempenho do modelo de regressão linear múltipla depende do atendimento das suposições de homoscedasticidade, de normalidade e de independência dos erros e da ausência de multicolinearidade. Sobre a última, Corrar (2007) comenta que o método *stepwise* funciona como ação corretiva nos casos de multicolinearidade, já que se duas variáveis são muito correlacionadas, estando uma delas no modelo, a outra tende a ser excluída por não acrescentar informação que explique a variável resposta, mantendo-se a de maior significância.

3.2.2 Qualidade do modelo

Para que se possa mensurar a qualidade do modelo proposto, são utilizadas as seguintes medidas: i) teste de Kolmogorov-Smirnov (KS) para duas amostras e ii) taxa de acerto de classificações nos grupos bom e mau. O primeiro calcula a máxima diferença entre as distribuições acumuladas de bons e maus pagadores e o segundo é obtido através da divisão do número de clientes classificados corretamente (soma da contagem de clientes bons classificados como bons e maus classificados como maus) pelo número total de observações da amostra.

Picinini *et al.* (2003) considera que, nos modelos de classificação obtidos através de regressão logística, valores de KS e a taxa de acertos superiores a 30% e 65%, respectivamente, são suficientes para considerar o modelo eficiente na separação e na predição de bons e maus pagadores. Porém, por não se tratar de um modelo de classificação e sim preditivo, estes valores podem servir como referência, não havendo especificação de valores ótimos.

3.3 SISTEMÁTICA PARA A ATRIBUIÇÃO DO LIMITE

A proposta para este estudo é a atribuição do limite através do lucro previsto. Para isso, utiliza-se o Lucro Observado, já definido na seção 3.1.1, dimensionando o valor total de vendas em termos da margem de lucro. O método para definir o limite de cada cliente consiste de mais três passos: o primeiro é redimensionar a variável resposta do modelo, o lucro previsto, para seu valor de venda correspondente, através da Equação 4, que equivale ao valor que o cliente deve adquirir em compras na empresa para que se observe o lucro previsto.

$$\text{Lucro Previsto} = \text{Valor de Venda} * \text{Margem de Lucro} \quad (4)$$

O segundo é definir um valor de limite para cada faixa de valor de venda e, após, classificar este valor na classe equivalente, com valores padronizados para o limite. Dessa forma, valores negativos de Lucro Previsto resultantes do modelo são automaticamente classificados no grupo de maus clientes e rejeitados. Portanto, apenas os clientes classificados no grupo de bons clientes, ou seja, com valor previsto positivo, terão algum limite atribuído.

4 RESULTADOS

Nesta seção, são apresentados e discutidos os resultados obtidos em cada etapa da construção, desde o pré-processamento até a avaliação da qualidade do modelo proposto a partir da utilização de dados reais e sua utilização para atribuição do limite de crédito.

4.1 PRÉ-PROCESSAMENTO

Esta etapa consiste da apresentação dos resultados da delimitação da população, da seleção da amostra e da análise preliminar das observações.

4.1.1 Delimitação da população

O banco de dados utilizado neste trabalho é composto por informações cadastrais e variáveis referentes ao desempenho perante a utilização do crédito de 17.005 clientes de uma rede de farmácias, com unidades distribuídas em várias cidades do Rio Grande do Sul, que oferece, como uma das formas de pagamento e parcelamento, o cartão próprio (*private label*). Todos estes clientes tiveram crédito concedido sem nenhum modelo implantado na empresa, sendo o único restritivo para negativa da concessão a presença do solicitante nos serviços de negativação, como SPC e SERASA.

No momento da consolidação dos registros, os clientes que tiveram crédito concedido classificaram-se quanto a sua utilização e desempenho de pagamento conforme apresentado na Tabela 2. Da categoria “A vencer”, fazem parte aqueles clientes que utilizaram o crédito da empresa, porém ainda não possuem nenhuma prestação vencida. “Não utilizado” são os clientes que nunca fizeram uso do crédito após a concessão e “Vencido” os que possuem pelo menos uma prestação em atraso. As categorias “Pagou adiantado”, “Pagou em atraso” e “Pagou no

dia” são compostas pelos clientes que pagaram a última parcela antes do vencimento, após o vencimento e no dia do vencimento, respectivamente.

Tabela 2. Distribuição dos clientes quanto à utilização do crédito

Classificação	Frequência
A vencer	1.206
Não utilizado	3.323
Pagou adiantado	3.301
Pagou em atraso	2.987
Pagou no dia	874
Vencido	5.314
Total	17.005

A primeira etapa é a identificação dos clientes perante a utilização do crédito e quanto aos critérios de lucro ou prejuízo. A empresa estima que sua margem de lucro sobre o valor de venda dos produtos é de 30%. Para que se possa classificar os clientes como bons ou maus, baseada na estimativa de margem de lucro fornecida pela empresa, é aplicada a Equação 5, substituindo o termo Margem de Lucro pela estimativa, obtendo

$$\text{Lucro Observado} = 0,3 * \text{Valor Pago} - \text{Valor Não Pago} \quad (5)$$

onde, por definição do concessor, Valor Pago representa o valor total pago em dia ou pago com até 60 dias de atraso e Valor Não Pago o valor total em aberto há mais de 60 dias ou pago após 60 dias de atraso, considerando que estes pagamentos apenas equilibram os gastos com cobrança e negativação decorrentes da inadimplência.

Dessa forma, clientes com Lucro Observado maior que zero compõe a categoria “bom” e menor ou igual que zero, a categoria “mau”. Além destes, também são classificados os clientes sem uso, que nunca utilizaram o crédito e os clientes indefinidos, que possuem apenas parcelas a vencer.

4.1.2 Seleção da amostra

As variáveis disponíveis para o estudo foram informadas pelos clientes ao preencher o formulário de proposta, no momento da solicitação do crédito e, após, inseridas no sistema da empresa. Tais observações foram coletadas para os clientes com tempo de cadastro entre 12 e

18 meses, para que os dados contemplassem um número suficientemente grande de clientes com situações de utilização de crédito e pagamento definidas.

Toda informação utilizada no modelo proposto é oriunda exclusivamente dos registros que a compõem, estando, assim, limitadas à disponibilidade dos registros. As variáveis identificadas na proposta de crédito como possíveis preditoras do lucro são apresentadas na Figura 2.

Variável	Descrição
Sexo	Feminino ou masculino
Naturalidade	Cidade de nascimento
Estado Civil	Casado, divorciado, solteiro, viúvo, etc.
Idade	Idade (em anos) no dia do cadastro
Escolaridade	Fundamental, médio ou superior
Tipo de Residência	Alugada, cedida, mora com os pais ou própria
CEP Residencial	CEP da rua de residência
Profissão	Profissão ou cargo exercido no dia do cadastro
Tipo de Ocupação	Assalariado, aposentado, autônomo, profissional liberal, etc.
Tempo de Serviço	Tempo (em meses) no emprego no dia do cadastro
Tipo de Renda	Comprovada ou declarada
CEP Comercial	CEP da rua de trabalho
Crédito 3ºs	Possui crédito em outros estabelecimentos
Pensão	Paga pensão
Filho	Tem filho

Figura 2. Variáveis identificadas na proposta de crédito

Como o interesse é trabalhar apenas com os clientes que já possuam histórico de compra e fechamento de situação de pagamento, apenas os grupos bom e mau serão utilizados, por possuírem informações que separam de forma mais precisa os perfis dos clientes quanto ao seu comportamento transacional, restando, assim, 12.476 clientes a serem considerados nas análises e construção do modelo. A distribuição do total de clientes da amostra pela classificação por sua utilização do crédito é apresentada na Tabela 3.

Tabela 3. Distribuição de clientes quanto à classificação do uso do crédito

Classificação	Frequência
Sem uso	3.323
Bom	7.694
Mau	4.782
Indefinido	1.206
Total	17.005

Antes de seguir com a análise, deve-se verificar a qualidade dos dados quanto ao seu preenchimento, padronização e consistência das informações. Por exemplo, a variável Naturalidade, por ser um campo de digitação aberto, possui várias escritas para a mesma cidade. O mesmo acontece na variável Profissão. Nestes casos, é necessária a padronização da escrita. Idades e tempo de serviço negativos, CEPs gerais ou zerados foram considerados *missings*, não sendo utilizados no estudo.

Após o tratamento dos dados, as observações são separadas aleatoriamente em duas amostras proporcionais, uma com 80% de todos os casos do grupo bom e do grupo mau, que constitui a amostra de análise, utilizada para a criação do modelo, e os demais 20% utilizados para posteriormente testar o poder de predição do modelo para observações provenientes de uma mesma população.

A amostra de análise fica, então, constituída de 9.981 observações, sendo 3.826 do grupo mau e 6.155 do bom e a de teste com 956 clientes do grupo mau e 1.539 do grupo bom, totalizando 2.495 observações.

4.1.3 Análise preliminar

Nesta etapa, para determinar as possíveis variáveis preditoras, é calculado o risco relativo (RR) para cada nível de cada atributo, dividindo-se o percentual de clientes do grupo bom pelo percentual do grupo mau de todos os níveis de cada variável e após, agrupados numa escala com 7 níveis, do péssimo ao excelente. Ambos os procedimentos foram mostrados na seção 3.1.3. Os atributos agrupados na classe de RR neutro não serão considerados para a construção do modelo, pois têm baixo poder de discriminação entre bons e maus clientes. Em razão disso, as variáveis Tipo de Renda, Crédito 3ºs e Filho foram descartadas da análise, pois suas classes de respostas possíveis obtiveram RR neutro. A variável Pensão também foi descartada pelo fato de que apenas 195 dos 9981 clientes da amostra de análise tiveram resposta positiva, portanto sua representatividade para a população é muito baixa.

Para trabalhar com as diferentes variáveis cadastrais, elas foram transformadas de modo a estarem aptas a ingressarem no modelo, utilizando a técnica de criação de variáveis *dummy*, que só assumem os valores um, quando possui a característica, ou zero, quando não possui. O número de variáveis *dummy* para cada variável original depende da forma que cada categoria é classificada.

Na variável CEP residencial, devido ao grande número de atributos, primeiramente os atributos que possuem menos de 30 observações são desconsiderados, pela baixa representatividade no conjunto de dados. Após, as *dummies* são criadas com base nos agrupamentos do risco relativo. Dessa forma, são definidos os CEPs conforme sua classificação de risco relativo entre péssimo e excelente, e depois criadas as *dummies* para cada classe. O mesmo é realizado para as variáveis CEP comercial, Naturalidade e Profissão. Os resultados para os agrupamentos e criação das *dummies* das respectivas variáveis são apresentados na Tabela 5.

Tabela 4. Criação de variáveis *dummy* por classe de risco relativo

Variável	Dummy	Mau	Bom	Total	RR	Classe de Risco
Grupos de profissões	DGPROF1	45	31	76	0,42	Péssimo
	DGPROF2	220	211	431	0,59	Muito Mau
	DGPROF3	1480	1893	3373	0,78	Mau
	DGPROF4	570	941	1511	1,01	Neutro
	DGPROF5	198	431	629	1,33	Bom
	DGPROF6	452	1373	1825	1,86	Muito Bom
	DGPROF7	58	201	259	2,12	Excelente
	Não classificado	384	495	879	-	-
	Não informado	419	579	998	-	-
Grupos de naturalidades	DCIDNA1	19	14	33	0,49	Péssimo
	DCIDNA2	84	81	165	0,64	Muito Mau
	DCIDNA3	1802	2063	3865	0,76	Mau
	DCIDNA4	257	408	665	1,05	Neutro
	DCIDNA5	498	974	1472	1,29	Bom
	DCIDNA6	230	627	857	1,80	Muito Bom
	DCIDNA7	92	337	429	2,43	Excelente
	Não classificado	567	1125	1692	-	-
	Não informado	277	526	803	-	-
Grupos de CEPs residenciais	DGCEPRE2	17	18	35	0,66	Muito Mau
	DGCEPRE3	2174	2944	5118	0,84	Mau
	DGCEPRE4	877	1393	2270	0,99	Neutro
	DGCEPRE5	471	930	1401	1,23	Bom
	DGCEPRE6	153	382	535	1,55	Muito Bom
	DGCEPRE7	119	438	557	2,29	Excelente
	Não classificado	11	45	56	-	-
		Não informado	4	5	9	-
Grupos de CEPs comerciais	DGCEPCO2	40	39	79	0,61	Muito Mau
	DGCEPCO3	1301	1702	3003	0,81	Mau
	DGCEPCO5	484	928	1412	1,19	Bom
	DGCEPCO6	113	302	415	1,66	Muito Bom
	DGCEPCO7	51	184	235	2,24	Excelente
	Não classificado	104	204	308	-	-
		Não informado	1733	2796	4529	-

Já na variável Estado Civil, por exemplo, é criada uma *dummy* para cada classe, conforme Tabela 4. As duas variáveis numéricas, Idade e Tempo de Serviço, foram categorizadas em faixas de valor, que também tiveram o risco relativo calculado e foram agrupadas conforme as classes de RR para a criação das *dummies*.

Tabela 5. Exemplo da criação de variáveis *dummy* para as variáveis categóricas

Estado civil	<i>Dummy</i>	Mau	Bom	Total	RR	Classe de Risco
Casado	DCASADO	935	2122	3057	1,41	Bom
Concubinado	DCONCUBINADO	29	44	73	0,94	Neutro
Divorciado	DDIVORCIADO	211	428	639	1,26	Bom
Separado	DSEPARADO	87	160	247	1,14	Bom
Solteiro	DSOLTEIRO	2105	2417	4522	0,71	Mau
Viúvo	DVIUVO	238	634	872	1,65	Muito Bom
Outros	DOUTROS	221	350	571	0,98	Neutro

Dessa forma, foram criadas 40 variáveis *dummy* em função das classes dos atributos e 28 em função do risco relativo, totalizando, assim, 68 variáveis *dummy* que passam a ser a totalidade das variáveis explicativas candidatas a ingressar no modelo.

4.2 CONSTRUÇÃO E AVALIAÇÃO DO MODELO

Esta etapa consiste da apresentação dos resultados de cada etapa da sistemática para a construção e avaliação da qualidade do modelo proposto.

4.2.1 Construção do modelo

O *software* utilizado para a construção do modelo de previsão do lucro através da regressão linear múltipla foi *IBM SPSS Statistics 20*.

Na seleção de variáveis, foi utilizado o método automático *stepwise*, que traz, no modelo final, apenas as variáveis que mais são significativas e influenciam o resultado da variável dependente. Além disso, este método tende a funcionar como ação corretiva nos casos de multicolinearidade.

Antes de se obter o modelo final, foi necessário realizar dois agrupamentos para que os sentidos das estimativas estivessem de acordo com o que seu risco relativo havia indicado

previamente, criando então, as variáveis DGCEPRE567 e DIDAD123, que juntam os grupos de CEP residenciais com risco relativo bom, muito bom e excelente e os grupos de idade de 18 a 30 anos, respectivamente. A Equação 6 apresenta o modelo encontrado.

$$\begin{aligned}
 y = & -10,43 + 38,076 DCASADO - 99,661 DCIDNA1 - 45,758 DCIDNA2 \\
 & - 25,974 DCIDNA3 + 27,505 DCIDNA7 - 31,628 DGPROF2 \\
 & + 22,954 DGPROF6 + 43,513 DGPROF7 - 33,378 DIDAD123 \\
 & - 22,015 DIDAD4 + 37,510DIDAD7 + 68,551 DIDAD8 \\
 & - 27,409 DRESALU - 37,853 DSEXOM + 34,643 DTSERV7 \\
 & + 51,040 DTSERV8 + 74,310 DTSERV9 + 33726 DGCEPRE567
 \end{aligned} \tag{6}$$

O método de seleção automática *stepwise* selecionou 18 variáveis significativas para explicar o lucro. A Figura 3 apresenta as variáveis *dummy* e as respectivas especificações.

DCASADO = É casado (a)	DIDAD123 = Idade entre 18 e 30 anos
DCIDNA1 = Cidade de nascimento com desempenho péssimo	DIDAD4 = Idade entre 31 e 35 anos
DCIDNA2 = Cidade de nascimento com desempenho muito mau	DIDAD7 = Idade entre 51 e 60 anos
DCIDNA3 = Cidade de nascimento com desempenho mau	DIDAD8 = Idade acima de 60 anos
DCIDNA7 = Cidade de nascimento com desempenho excelente	DRES_ALU = Mora em residência alugada
DGCEPRE567 = CEP residencial com desempenho mau, muito mau ou péssimo	DSEXOM = É do sexo masculino
DGPROF2 = Profissão com desempenho muito mau	DTSERV7 = Tempo de serviço entre 61 e 90 meses
DGPROF6 = Profissão com desempenho muito bom	DTSERV8 = Tempo de serviço entre 91 e 120 meses
DGPROF7 = Profissão com desempenho excelente	DTSERV9 = Tempo de serviço maior de 120 meses

Figura 3. Variáveis *dummy* utilizadas no modelo

Após a definição do modelo final, deve-se verificar se este atende às suposições da regressão linear múltipla. A primeira a ser investigada é a de homoscedasticidade, em que busca-se a variância constante dos erros. Através do gráfico de dispersão entre os valores preditos e os resíduos, mostrado na Figura 4, pode-se observar que os erros se distribuem aleatoriamente em torno de zero, atendendo então, à primeira suposição.

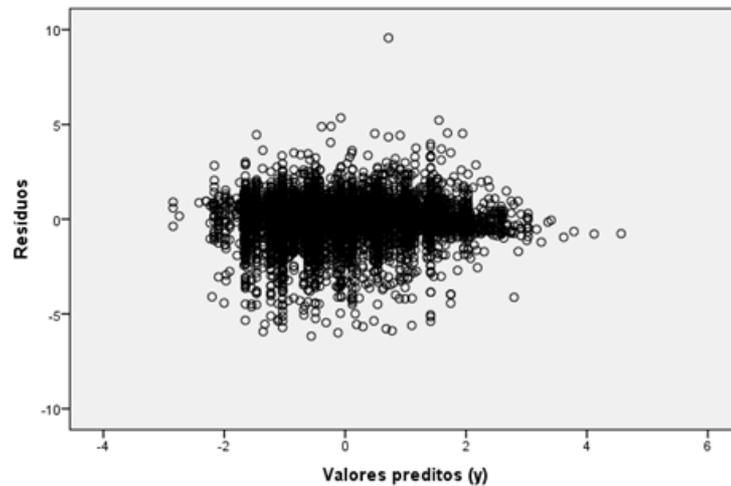


Figura 4. Distribuição dos valores preditos pelos resíduos

A normalidade dos erros pode ser verificada através do histograma apresentado na Figura 5, que aparenta se adequar de forma aproximada à curva normal, atendendo assim a mais uma das suposições.

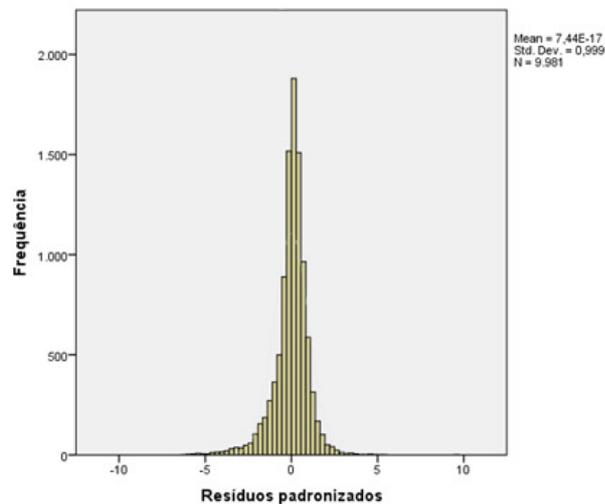


Figura 5. Histograma dos resíduos padronizados

Para verificar a independência dos erros, utiliza-se o teste de Durbin-Watson, cuja estatística de teste encontrada foi d_w igual a 1,991. Conforme a tabela de valores críticos para este teste, $d_U = 1,7$ e $d_U < 1,991 \leq 4 - d_U$, portanto, esta suposição também está atendida. A multicolinearidade entre as variáveis foi reduzida utilizando o método *stepwise*, mas também pode ser observada através dos VIF, que foram todos menores que 2, enquanto o valor máximo aceitável se aproxima de 10. Dessa forma, estas duas suposições também estão validadas.

4.2.2 Qualidade do modelo

Para mensurar o desempenho e a qualidade do modelo final, são utilizadas a taxa de acerto de classificações nos grupos bom e mau e o valor do teste de Kolmogorov-Smirnov (KS) para duas amostras.

As classificações original e predita pelo modelo nas amostras de análise e teste são apresentadas na Tabela 6. Na amostra de análise, 6.057 das 9.981 observações foram corretamente classificadas, obtendo-se o percentual de classificação correta de 60,59%. Na amostra de teste, este valor é de 60,08%, com um total de 1.499 acertos em 2.495 observações. Estes valores mostram que o modelo tem desempenho homogêneo para os clientes já existentes (utilizados na construção do modelo) e para os futuros proponentes a crédito deste negócio.

Tabela 6. Quantidades e percentuais de acertos do modelo

Amostra	Classificação original	Classificação predita		
		Mau	Bom	Total
Análise	Mau	2766 (72,3%)	1060 (27,7%)	3826
	Bom	2864 (46,5%)	3291 (53,5%)	6155
	Total	5630 (56,4%)	4351 (43,6%)	9981
Teste	Mau	683 (71,4%)	273 (28,6%)	956
	Bom	723 (47,0 %)	816 (53,0%)	1539
	Total	1406 (56,4%)	1089 (43,6%)	2495
Total	Mau	3449 (72,1%)	1333 (27,9%)	4782
	Bom	3587 (46,6%)	4107 (53,4%)	7694
	Total	7036 (56,4%)	5440 (43,6%)	12476

Observando os acertos pela classificação original, percebe-se que o desempenho do modelo está mais favorável a não conceder crédito aos maus pagadores, acertando mais nos casos provenientes do grupo mau (em torno de 70%), porém negando a concessão para quase metade dos clientes lucrativos.

Através do resultado do teste não paramétrico de Kolmogorov Smirnov (KS), pode-se determinar se duas amostras provêm de populações distintas, o que significaria que o modelo consegue separar os grupos bom e mau. A Figura 6 mostra o grau de separação entre as duas amostras.

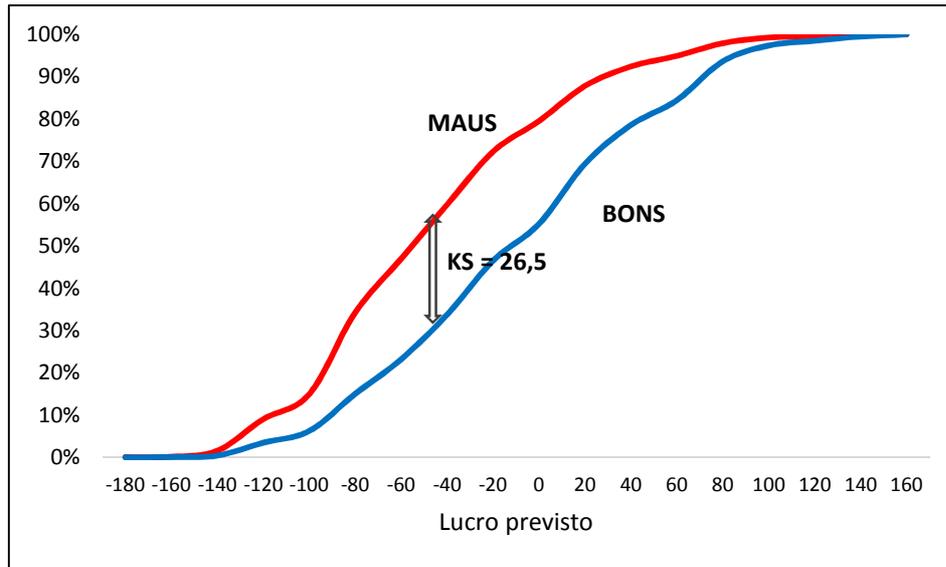


Figura 6. Representação do valor de KS para o modelo construído

O valor resultante do teste KS é de 26,5, o que significa que o modelo consegue separar razoavelmente bem os dois grupos. A fim de avaliar os resultados monetários obtidos pelo modelo, as medidas descritivas da variável de lucro são comparadas no cenário sem a utilização de nenhum modelo de risco de crédito e com a utilização do modelo proposto. O primeiro é apresentado na Tabela 7.

Tabela 7. Medidas de lucro para os grupos de clientes observados

Amostra	Tipo	N	Média (R\$)	Desvio padrão (R\$)	Soma (R\$)
Análise	Mau	3.826	-197,88	210,07	-757.105,86
	Bom	6.155	113,34	121,81	697.584,11
	Total	9.981	-5,96	221,27	-59.521,75
Teste	Mau	956	-209,07	224,16	-199.875,60
	Bom	1.539	114,33	125,16	175.949,27
	Total	2.495	-9,59	231,58	-23.926,33
Total	Mau	4.782	-200,12	212,98	-956.981,46
	Bom	7.694	113,53	122,48	873.533,38
	Total	12.476	-6,69	223,37	-83.448,08

No cenário anterior à implementação do modelo de crédito, a empresa aprovou a concessão de crédito para todos os proponentes que não possuíam algum restritivo externo, como SPC e SERASA e, desta forma, aprovou 17.005 clientes, dos quais 12.476 possuíam comportamento de utilização de crédito consolidado. Destes, 4.782 resultaram em prejuízo e 7.694 em lucro, contabilizando um prejuízo médio por cliente de R\$ 6,69, perdendo em média R\$ 200,12 com cada cliente do grupo mau e ganhando em média R\$ 113,53 com cada cliente

do grupo bom. O valor total resultou num prejuízo de R\$ 83.448,08. Considerando a implementação do modelo de risco de crédito proposto, os resultados são os que seguem apresentados na Tabela 8.

Tabela 8. Medidas de lucro para os grupos de clientes previstos pelo modelo

Amostra	Decisão	N	Média (R\$)	Desvio padrão (R\$)	Soma (R\$)
Análise	Negar	5630	-51,96	229,20	-292.553,06
	Aprovar	4351	53,56	195,11	233.031,31
	Total	9981	-5,96	221,27	-59.521,75
Teste	Negar	1406	-54,36	244,34	-76.434,99
	Aprovar	1089	48,22	199,76	52.508,66
	Total	2495	-9,59	231,58	-23.926,33
Total	Negar	7036	-52,44	232,29	-368.988,06
	Aprovar	5440	52,49	196,04	285.539,98
	Total	12476	-6,69	223,37	-83.448,08

No cenário apresentado na Tabela 8, o total de 5.440 proponentes teriam seu crédito concedido, que corresponde a 43,59% de aprovação, os quais resultariam em um ganho médio de R\$ 52,49 por cliente, obtendo um lucro total de R\$ 285.539,08.

Para avaliar e validar o modelo, verifica-se a amostra de análise quanto à distribuição dos grupos de cliente pelo lucro esperado determinado pelo modelo, além das taxas de sinistro e aprovação, como apresentado na Tabela 9.

Tabela 9. Distribuição dos clientes e taxas de sinistro e de aprovação do modelo

Lucro previsto (R\$)	Mau			Bom			Total			Taxa de sinistro (%)		Taxa de aprovação (%)
	#	%	% acum.	#	%	% acum.	#	%	% acum.	Classe	Total	
até -160	3	0,08	0,08	1	0,02	0,02	4	0,04	0,04	75,00	38,33	100,00
de -160 a -140	3	0,08	0,16	6	0,10	0,11	9	0,09	0,13	33,33	38,32	99,96
de -140 a -120	55	1,44	1,59	22	0,36	0,47	77	0,77	0,90	71,43	38,32	99,87
de -120 a -100	286	7,48	9,07	186	3,02	3,49	472	4,73	5,63	60,59	38,06	99,10
de -100 a -80	226	5,91	14,98	171	2,78	6,27	397	3,98	9,61	56,93	36,94	94,37
de -80 a -60	738	19,29	34,27	539	8,76	15,03	1277	12,79	22,40	57,79	36,06	90,39
de -60 a -40	487	12,73	46,99	504	8,19	23,22	991	9,93	32,33	49,14	32,47	77,60
de -40 a -20	489	12,78	59,78	649	10,54	33,76	1138	11,40	43,73	42,97	30,03	67,67
de -20 a 0	479	12,52	72,29	786	12,77	46,53	1265	12,67	56,41	37,87	27,40	56,27
de 0 a 20	278	7,27	79,56	535	8,69	55,22	813	8,15	64,55	34,19	24,36	43,59
de 20 a 40	316	8,26	87,82	874	14,20	69,42	1190	11,92	76,48	26,55	22,10	35,45
de 40 a 60	176	4,60	92,42	564	9,16	78,59	740	7,41	83,89	23,78	19,85	23,52
de 60 a 80	97	2,54	94,96	362	5,88	84,47	459	4,60	88,49	21,13	18,03	16,11
de 80 a 100	114	2,98	97,94	565	9,18	93,65	679	6,80	95,29	16,79	16,80	11,51
de 100 a 120	50	1,31	99,24	231	3,75	97,40	281	2,82	98,11	17,79	16,81	4,71
de 120 a 140	15	0,39	99,63	69	1,12	98,52	84	0,84	98,95	17,86	15,34	1,89
de 140 a 160	8	0,21	99,84	62	1,01	99,53	70	0,70	99,65	11,43	13,33	1,05
mais de 160	6	0,16	100,00	29	0,47	100,00	35	0,35	100,00	17,14	17,14	0,35
TOTAL	3826	100	-	6155	100	-	9981	100	-	-	-	-

A taxa de sinistro por classe mostra quantos maus pagadores seriam aprovados em cada faixa de lucro previsto pelo modelo utilizando como ponto de corte o limite inferior da classe, enquanto a taxa de sinistro total mostra o número de maus pagadores sobre o total de aprovados acumulado por classe. A taxa de aprovação representa o total de clientes aprovados acumulado em cada faixa de lucro previsto.

Utilizando como ponto de corte o lucro previsto igual a zero, em que todos os proponentes com estimativa de prejuízo são negados, 43,59% das propostas seriam aprovadas e, destes, aproximadamente 24,36% seriam maus pagadores. Em relação aos grupos bom e mau, 62,3% do primeiro teriam a concessão negada e do segundo, 53,5% teriam crédito concedido. A Figura 7 mostra a distribuição dos clientes dos dois grupos nas faixas de lucro esperado e a taxa de sinistro.

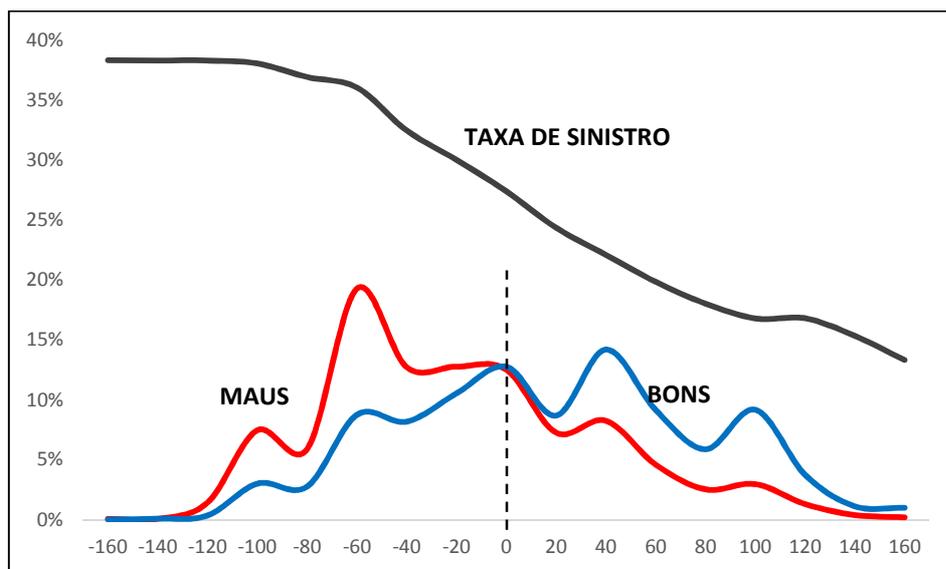


Figura 7. Distribuição dos clientes e taxa de sinistro com modelo de previsão

Como já verificado na Tabela 5, os clientes provenientes do grupo mau tendem a ser classificados corretamente, enquanto no grupo bom o modelo é menos assertivo. A taxa de sinistro decrescente também mostra que conforme o ponto de corte avança na direção dos maiores lucros esperados, a taxa de maus clientes tende a ser menor.

A comparação entre as faixas de lucro esperado e o lucro observado em ambos os grupos de clientes da amostra de análise é apresentada na Tabela 10.

Tabela 10. Análise comparativa do lucro esperado com lucro observado – amostra de análise

Lucro previsto (R\$)	Mau		Bom		Total		Aprovação acumulada	
	#	R\$	#	R\$	#	R\$	%	R\$
até -160	3	-456,13	1	10,87	4	-445,26	100,00	-59.521,75
de -160 a -140	3	-1.563,69	6	194,62	9	-1.369,07	99,96	-59.076,49
de -140 a -120	55	-14.297,91	22	2.525,09	77	-11.772,81	99,87	-57.707,42
de -120 a -100	286	-69.814,64	186	18.519,66	472	-51.294,98	99,10	-45.934,61
de -100 a -80	226	-59.587,24	171	18.585,61	397	-41.001,63	94,37	5.360,37
de -80 a -60	738	-164.918,85	539	55.834,54	1.277	-109.084,31	90,39	46.362,00
de -60 a -40	487	-93.871,53	504	52.623,13	991	-41.248,40	77,60	155.446,31
de -40 a -20	489	-101.003,06	649	70.199,06	1.138	-30.803,99	67,67	196.694,72
de -20 a 0	479	-84.100,06	786	78.567,45	1.265	-5.532,60	56,27	227.498,71
de 0 a 20	278	-51.922,47	535	59.734,29	813	7.811,81	43,59	233.031,31
de 20 a 40	316	-48.812,10	874	96.648,00	1.190	47.835,89	35,45	225.219,50
de 40 a 60	176	-28.688,81	564	73.445,48	740	44.756,67	23,52	177.383,61
de 60 a 80	97	-12.171,97	362	41.077,43	459	28.905,46	16,11	132.626,93
de 80 a 100	114	-16.687,82	565	78.541,93	679	61.854,11	11,51	103.721,48
de 100 a 120	50	-6.321,28	231	30.235,09	281	23.913,81	4,71	41.867,37
de 120 a 140	15	-1.640,26	69	7.992,76	84	6.352,50	1,89	17.953,56
de 140 a 160	8	-196,72	62	9.060,97	70	8.864,26	1,05	11.601,06
mais de 160	6	-1.051,32	29	3.788,13	35	2.736,80	0,35	2.736,80
TOTAL	3.826	-757.105,86	6.155	697.584,11	9.981	-59.521,75	-	-

Analisando os resultados apresentados, observa-se que os 3.826 clientes do grupo mau geraram um prejuízo de R\$ 757.105,86, enquanto os 6.155 do grupo bom resultaram num ganho de R\$ 697.584,11. Ao todo, estes clientes geraram um prejuízo de R\$ 59.521,75. O lucro máximo é obtido na primeira faixa de lucro previsto positivo. Definindo o limite inferior desta classe como ponto de corte, a taxa de aprovação seria de 43,59%, ou seja, menos da metade das propostas seriam aprovadas, porém, o valor final contabilizaria um lucro de R\$ 233.031,31.

A empresa poderia, ainda, utilizar como ponto de corte um valor de lucro previsto negativo, como por exemplo, de -R\$ 100,00, pois neste caso, mesmo aprovando clientes com os quais espera-se prejuízo, no valor acumulado o resultado ainda seria positivo. Nesse caso, a taxa de aprovação seria de 94,37% e o lucro de R\$ 5.360,37. Qualquer ponto de corte acima deste valor resultaria em lucro para a companhia, mesmo aumentando a taxa de aprovação de maus clientes.

Através da Figura 8, que compara o resultado geral com os lucros e prejuízos esperados, também mostra que o lucro com os bons clientes supera o prejuízo com os maus a partir dos valores de lucro esperado maiores que -R\$ 100,00 e que o maior valor de lucro é observado na primeira faixa de lucro previsto positivo, o que corrobora a definição do ponto de corte nesta classe.

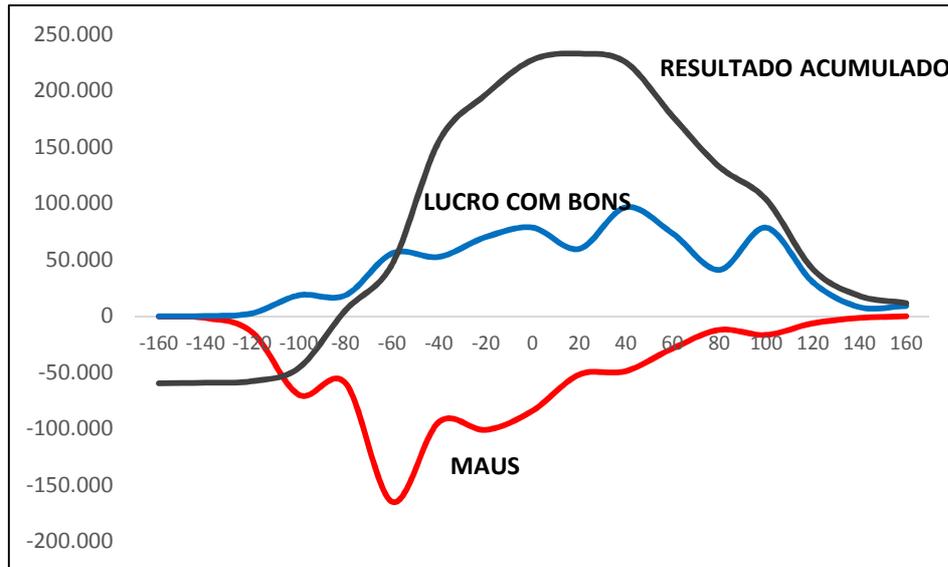


Figura 8. Prejuízos e lucro em cada grupo e total com a aprovação de crédito

A Tabela 11 apresenta um resumo da comparação dos resultados do cenário que aprova todos os proponentes que solicitam crédito à empresa com o que utiliza o modelo proposto, mostrando o valor médio de lucro esperado por cliente aprovado e negado pelo modelo de previsão, considerando as observações dos grupos bom e mau de ambas as amostras, de análise e de teste.

Tabela 11. Comparativo dos cenários sem e com modelo de previsão de crédito

Lucro previsto	Sem modelo	Com modelo	
		Aprovados	Negados
Cientes aprovados	12.476	5.440	7.036
Média (R\$)	-6,69	52,49	-52,44
Total (R\$)	-83.448,08	285.539,98	-368.988,06

Pode-se observar que a implementação do modelo de previsão do lucro inverte o quadro de prejuízo observado na empresa na ordem de R\$ 83.448,08 para um lucro estimado de R\$ 285.539,98 através dos valores e desempenhos praticados, em que cada cliente negado pelo modelo representaria um prejuízo de R\$ 52,44 no resultado da empresa, se este fosse aprovado.

4.3 ATRIBUIÇÃO DO LIMITE

A atribuição do limite através do lucro previsto consiste do redimensionamento da variável resposta do modelo para seu valor de venda correspondente. Para isso, utiliza-se a Equação 4, apresentada na seção 3.3, substituindo o termo Margem de Lucro pela estimativa

da empresa de 30% e, após manipulação dos termos, obtendo a Equação 7, que representa o Valor de Venda em função do Lucro Previsto e da Margem de Lucro.

$$\text{Valor de Venda} = \frac{\text{Lucro Previsto}}{0,3} \quad (7)$$

Os valores obtidos da Equação 7 representam, em valor monetário, quanto o cliente deve adquirir em produtos para que se obtenha o lucro esperado através da margem de lucro estimada e, portanto, correspondem ao limite recomendado ao cliente. Para que se obtenham limites padronizados, estes valores devem ser classificados em classes. Inicialmente, a empresa concedeu 43 valores de limite diferentes, conforme apresentados na Tabela 12.

Dessa forma, as faixas de limites concedidos são muito variadas, pouco amplas e mal distribuídas. O método sugerido para a atribuição do limite em faixas melhor distribuídas e coerentes com o lucro esperado por cliente é o cálculo do valor médio de venda esperado por faixa de lucro previsto.

Tabela 12. Limites concedidos pela empresa antes da implantação do modelo

Limite concedido (R\$)	Frequência	Percentual (%)	Limite concedido (R\$)	Frequência	Percentual (%)
50,00	221	1,8%	260,00	122	1,0%
60,00	391	3,1%	270,00	97	0,8%
70,00	455	3,6%	280,00	71	0,6%
75,00	3	0,0%	290,00	79	0,6%
80,00	510	4,1%	300,00	182	1,5%
90,00	804	6,4%	310,00	72	0,6%
100,00	743	6,0%	320,00	58	0,5%
110,00	700	5,6%	330,00	82	0,7%
120,00	970	7,8%	340,00	48	0,4%
130,00	540	4,3%	350,00	83	0,7%
140,00	1000	8,0%	360,00	67	0,5%
150,00	2274	18,2%	370,00	34	0,3%
160,00	323	2,6%	380,00	47	0,4%
170,00	239	1,9%	390,00	57	0,5%
180,00	318	2,5%	400,00	719	5,8%
190,00	182	1,5%	430,00	1	0,0%
200,00	231	1,9%	450,00	5	0,0%
210,00	195	1,6%	500,00	6	0,0%
220,00	130	1,0%	600,00	4	0,0%
230,00	136	1,1%	800,00	1	0,0%
240,00	122	1,0%	1000,00	2	0,0%
250,00	152	1,2%			

A Tabela 13 mostra o limite médio e o total concedido pela empresa e os limites sugeridos pelo estudo em cada faixa de lucro previsto. Os resultados mostram que os limites médios atribuídos pela empresa variam muito pouco entre cada classe de lucro esperado.

Tabela 13. Limite atribuído e limite sugerido pelo modelo

Lucro previsto (R\$)	Clientes	Limite atribuído (R\$)		Limite sugerido (R\$)	
		Médio	Total	Médio	Total
de 0 a 20	1006	166,98	167.980,00	20,81	20.932,94
de 20 a 40	1499	164,25	246.210,00	94,41	141.519,90
de 40 a 60	909	178,59	162.340,00	173,87	158.047,54
de 60 a 80	569	173,55	98.750,00	221,22	125.876,17
de 80 a 100	857	172,66	147.970,00	286,97	245.932,27
de 100 a 120	359	174,29	62.570,00	377,38	135.478,14
de 120 a 140	106	196,42	20.820,00	427,39	45.303,19
de 140 a 160	92	208,59	19.190,00	495,61	45.596,53
mais de 160	43	206,28	8.870,00	604,50	25.993,51
TOTAL	5440	171,82	934.700,00	173,65	944.680,20

Os clientes com previsão de lucro de mais de R\$ 160,00 (última classe) têm aproximadamente apenas R\$ 40,00 a mais de limite que aqueles com previsão de lucro menor que R\$ 20,00 (primeira classe). Através do método sugerido para a atribuição do limite, cada faixa de lucro esperado passa a ter um limite médio condizente com a classe. Ainda, por questões de padronização para a empresa, estes valores podem ser arredondados conforme apresentado na Tabela 14.

Tabela 14. Limites sugeridos padronizados

Lucro previsto (R\$)	Limite sugerido (R\$)	
	Médio	Total
de 0 a 20	25,00	25.150,00
de 20 a 40	100,00	149.900,00
de 40 a 60	175,00	159.075,00
de 60 a 80	225,00	128.025,00
de 80 a 100	300,00	257.100,00
de 100 a 120	375,00	134.625,00
de 120 a 140	425,00	45.050,00
de 140 a 160	500,00	46.000,00
mais de 160	600,00	25.800,00
TOTAL	173,65	970.725,00

Todos os passos para a decisão da concessão de crédito, determinação do ponto de corte e atribuição do limite estão contempladas na Figura 9.

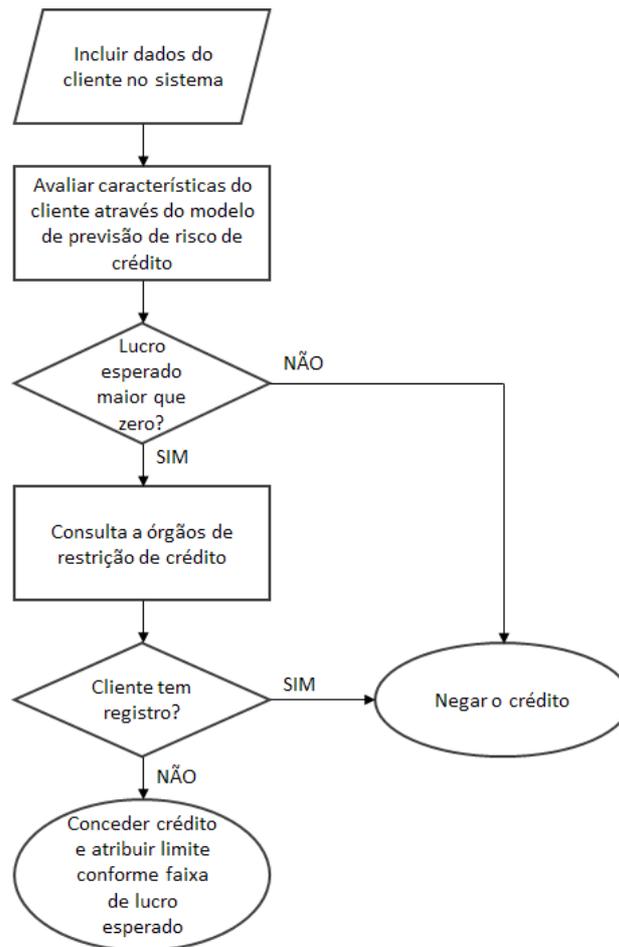


Figura 9. Fluxograma para decisão de crédito e atribuição e limite

Além de passar de um cenário de prejuízo para um de lucro, a empresa também economiza mais de 50% no que diz respeito ao valor das consultas nos órgãos de restrição de crédito, como SPC e Serasa, já que só serão consultados os clientes previamente aprovados pelo modelo que, como já visto, ao definir como ponto de corte os valores positivos para o lucro previsto, tem uma taxa de aprovação igual a 43,59%.

5 CONSIDERAÇÕES FINAIS

Num cenário de expansão do crédito, onde as classes mais baixas passam a ter maior participação no mercado, as ferramentas que auxiliam na tomada de decisão tornam-se cada vez mais necessárias e procuradas. Por estar exposto a uma fatia maior da população, o risco de inadimplência aumenta em grandes proporções.

Neste sentido, surge a proposta de utilizar uma ferramenta onde é possível prever, em medida monetária, o lucro que se espera com cada cliente, sendo ele um pagador fiel ou até

mesmo um que em dado momento não paga sua dívida, e através dessa resposta, atribuir um limite de crédito que corresponda em valor de venda de mercadorias, ao ganho previsto dada a margem de lucro.

De posse do modelo proposto, é realizada a análise dos resultados obtidos sem a utilização de nenhum modelo, em que todos os clientes que não possuíssem registro nos órgãos de restrição de crédito, como SPC e SERASA, tiveram crédito concedido, e comparado com os resultados obtidos com a utilização do modelo que prevê o lucro com cada cliente e atribui o limite adequado em relação ao seu lucro previsto. Com isso, têm-se um cenário em que, mesmo concedendo crédito para menos da metade dos clientes previamente aprovados, o resultado acumulado passa de prejuízo para lucro, utilizando como ponto de corte o valor mínimo de lucro esperado de R\$ 0,00, sendo o valor pago pelo cliente o suficiente para arcar com as despesas proporcionais de manutenção do negócio.

Sem a utilização de nenhum modelo, observou-se que, em média, cada cliente aprovado gerou uma perda de R\$ 6,69, enquanto que com a utilização do modelo proposto, espera-se um ganho médio de R\$ 52,49 por cliente, representando um aumento de aproximadamente 885% no resultado esperado com cada cliente.

A análise de sensibilidade mostra que o modelo também se mostra efetivo mesmo aprovando aqueles com os quais se espera prejuízo máximo de R\$ 100,00, o resultado acumulado continua sendo positivo. Nesse sentido, o modelo de previsão com variável resposta contínua, que avalia o lucro esperado, mostra-se uma ferramenta efetiva na concessão de crédito.

A utilização destes resultados, entretanto, se restringe ao caso particular em estudo, de concessão de crédito de empresa comercial que utiliza o crédito próprio como fidelização do cliente e impulso para as vendas. Aplicações em contextos diferentes devem ser adaptadas, pois o desempenho dos clientes quanto ao pagamento e, conseqüentemente, ao resultado monetário, variam nos diferentes ramos de comércio e serviço, principalmente no que diz respeito à margem de lucro praticada. Além disso, mesmo sendo aplicado na mesma empresa, o modelo deve ser revisado em períodos determinados para que acompanhe as possíveis mudanças de comportamento, dadas principalmente por fatores externos, como economia, taxa de desemprego, etc.

Apesar do bom desempenho do modelo proposto, algumas limitações podem ser melhor estudadas futuramente, como i) verificação dos resultados através da cobrança de multas por atraso e possíveis alterações dos conceitos de atrasos aceitáveis e inaceitáveis, considerando,

por exemplo, que atrasos de até 90 dias são aceitáveis pois, nesse contexto, há uma cobrança de juros que compensa o atraso; ii) utilização apenas de contratos encerrados, pois a maturação pode impactar no cálculo do lucro e iii) definição da margem de lucro diferente da estimada pela empresa, de 30%.

6 REFERÊNCIAS BIBLIOGRÁFICAS

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. *Análise multivariada: para cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2007.

DIEESE. *A evolução do crédito na economia brasileira (2008-2013)*. Nota técnica, n. 135, mai, 2014.

DINIZ, C. A. R.; LOUZADA NETO, F. Modelagem Estatística para Risco de Crédito. In: *Minicurso no XX SINAPE*, João Pessoa-PB, 2012.

GONÇALVES, E. B. *Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos*. Dissertação de Mestrado, Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2005.

GREENE, W. *Econometric Analysis*. 6th ed. New Jersey: Prentice Hall, 2003.

GUJARATI, D. N.; PORTER, D. C. *Basic Econometrics*. 5th ed. Boston: McGraw-Hill Irwin, 2009.

HAYKIN, S. *Redes neurais: princípios e práticas*. Trad. Paulo Martins Engel. 2ed. Porto Alegre: Bookman, 2001.

LAWRENCE, D. *Handbook of consumer lending*. New Jersey: Prentice Hall, 1992.

LEWIS, E. M. *An introduction to credit scoring*. San Rafael: Fair Isaac and Co., Inc, 1992.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. 4ed. Hoboken: Wiley-Interscience, 2006.

NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; LI, W. *Applied Linear Statistical Models*. 5ed. New York: McGraw-Hill, 2005.

PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. Mineração de critério de credit scoring utilizando algoritmos genéticos. In: *IV Simpósio Brasileiro de Automação Inteligente*, Bauru, SP, 2003.

SÁ, C. A. *Estabelecimento de limite de crédito: uma nova abordagem para um velho problema*. Rio de Janeiro: Qualitymark, 2004.

SELAU, L. P. R. *Modelagem para Concessão de Crédito a pessoas físicas em empresas comerciais: da decisão binária para a decisão monetária*. Tese de Doutorado, Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul, 2012.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete sample). *Biometrika*, Great Britain, v. 52, n. 3, p. 591-611, 1965.

SILVA, J. A. *Análise do estabelecimento do limite de crédito - Um estudo de caso*. Dissertação de Mestrado, Departamento de Economia, Contabilidade, Administração e Secretariado, Universidade de Taubaté, 2002.

STEINER, M. T. A.; CARNIERI, C.; KOPITTKKE, B. H.; STEINER NETO, P. J. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. *Revista de Administração*, São Paulo, v. 34, n. 3, p. 56-67, 1999.

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.