



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA, CONTABILIDADE E
SECRETARIADO EXECUTIVO
CURSO DE CIÊNCIAS ATUARIAIS

MACELLA VASCONCELOS DOS SANTOS

ANÁLISE DA SUSTENTABILIDADE DOS BENEFICÁRIOS DE UMA
OPERADORA DE PLANO DE SAÚDE: EVIDÊNCIAS PARA O NORTE E
NORDESTE

FORTALEZA

2017

MACELLA VASCONCELOS DOS SANTOS

**ANÁLISE DA SUSTENTABILIDADE DOS BENEFICÁRIOS DE UMA
OPERADORA DE PLANO DE SAÚDE: EVIDÊNCIAS PARA O NORTE E
NORDESTE**

Monografia apresentada à Faculdade de
Economia, Administração, Atuária,
Contabilidade e Secretariado, como requisito
parcial para obtenção do Grau de Bacharel em
Ciências Atuariais.

Orientador: Prof. Me. Luciana Moura
Reinaldo

Co-Orientador: Prof. Me. Rômulo Alves
Soares

FORTALEZA

2017

MACELLA VASCONCELOS DOS SANTOS

**ANÁLISE DA SUSTENTABILIDADE DOS BENEFICÁRIOS DE UMA
OPERADORA DE PLANO DE SAÚDE: EVIDÊNCIAS PARA O NORTE E
NORDESTE**

Monografia apresentada à Faculdade de Economia, Administração, Atuária, Contabilidade e Secretariado, como requisito parcial para obtenção do Grau de Bacharel em Ciências Atuariais.

Aprovada em : ____/____/____

BANCA EXAMINADORA

Prof. Me. Luciana Moura Reinaldo
Professor Orientador

Prof. Me. Rômulo Alves Soares
Professor Co-orientador

Prof. Me. Sérgio Cardoso
Membro da Banca Examinadora

À Deus, pelo dom da vida.

AGRADECIMENTOS

Agradeço a Deus, por todas as oportunidades que tive nesta caminhada.

Aos meus pais, em especial minha mãe, Márcia Maria, por todo apoio e dedicação. Aos meus irmãos Andréa e Pedro pela cumplicidade. Aos meus primos Amanda, Daniel e Thais, pela irmandade. Essas pessoas que sempre acreditaram em mim.

As minhas amigas de colégio, aos meus amigos de Maranguape e meus amigos de faculdade, que sempre me incentivaram e apoiaram.

Agradeço aos meus colegas de trabalho, em especial o meu gestor Marx Fiuza, por toda aprendizagem, incentivo e apoio nesta jornada.

Ao Emilson Aguiar, por toda ajuda durante os anos de faculdade. O professor Rômulo Alves, que foi um excelente colega de faculdade e hoje um ótimo professor e me ajudou muito na realização desta monografia. A minha orientadora Lucina Reinaldo, por todo o aprendizado e que sempre acreditou em mim.

Agradeço a todos que me ajudaram chegar até aqui.

"O homem não é nada além daquilo que a educação faz dele."

(Immanuel Kant)

RESUMO

As Operadoras de Planos de Saúde (OPS) cresceram no cenário econômico brasileiro, mesmo com algumas dificuldades impostas pelas normas regulativas da ANS que as impedem de negociar livremente, em termos de precificação, seus planos de acordo com as características de cada mercado. Desta forma, a análise de solvência de suas carteiras tornaram-se cada vez mais importantes para um mercado altamente competitivo. Assim este trabalho buscou identificar a melhor técnica estatística para uma análise de solvência. A pesquisa teve como amostra cerca de 3.423 beneficiários de uma OPS atuante no Norte e Nordeste. Foram utilizadas quatro técnicas para prever a sustentabilidade da carteira: regressão logística, análise discriminante, árvore de classificação e regressão, e *random forest*. Para analisar a técnica com melhor performance foi utilizado a Curva de Roc. Os resultados evidenciaram que a técnica estatística é a regressão logística é que melhor identifica as variáveis que explicam a sustentabilidade de uma carteira de plano de saúde.

Palavras-chave: Saúde Suplementar. Regressão Logística. Sustentabilidade.

ABSTRACT

The Operators of Health Plans (OPS) grew in the Brazilian economic scenario. But they are barred by ANS regulatory rules that prevent them from freely negotiating, in terms of pricing, their plans according to the characteristics of each market. In this way the solvency analysis of your portfolios have become increasingly important for a highly competitive market. Thus, this work sought to identify the best statistical technique for a solvency analysis. The survey sampled about 3.423 beneficiaries of an OPS in the state of North and Northeast. Four techniques were used to predict the sustainability of the portfolio: logistic regression, discriminant analysis, tree classification and regression, and random forest. To analyze the technique with better performance, the Roc Curve was used. The results showed that the statistical technique of logistic regression best identifies the variables that explain the sustainability of a health insurance portfolio.

Keywords: Supplementary Health. Logistic Regression. Sustainability.

LISTA DE FIGURAS

Tabela 1 – Distribuição dos beneficiários por estado de insustentabilidade e sexo	28
Tabela 2 – Distribuição dos beneficiários por estado de insustentabilidade e cidade	28
Tabela 3 – Estatísticas descritivas da idade, custo e receita dos beneficiários por grupo	29
Tabela 4 – Classificação dada pela regressão logística.....	31
Tabela 5 – Classificação dada pela análise discriminante.....	31
Tabela 6 – Classificação dada pela árvore de classificação e regressão	31
Tabela 7 – Classificação dada pelo Random Forest	32
Tabela 8 – Área abaixo da curva para as quatro técnicas empregadas	33

LISTA DE ILUSTRAÇÕES

Figura 1 – Histórico dos índices inflacionários do setor de saúde suplementar.....	17
Figura 2 – Histograma com a variável Custo original e transformada pela raiz quadrada	30
Figura 3 – Histograma com a variável Receita original e transformada pela raiz quadrada....	30
Figura 4 – Curvas ROC para as quatro técnicas aplicadas.....	33

LISTA DE QUADROS

Quadro 1 – Variáveis utilizadas no estudo.....	23
--	----

SUMÁRIO

1	INTRODUÇÃO	12
2	REFERENCIAL TEÓRICO	14
2.1	Mercado brasileiro de saúde suplementar	14
2.2	Estudos sobre previsão de insustentabilidade de beneficiários de planos de saúde	19
3	METODOLOGIA	22
3.1	Tipologia da Pesquisa	22
3.2	População e amostra.....	22
4	RESULTADOS	28
4.1	Análise descritiva e inferencial.....	28
4.2	Análise multivariada	30
5	CONCLUSÃO	34

1 INTRODUÇÃO

A Constituição Federal de 1988 definiu como direito social e obrigação do Estado a prestação de serviços de saúde. Para integralizar todos os preceitos do art.196 da CF/88 em um sistema. Com a Lei 8.080, foi criado o Sistema Único de Saúde (SUS).

A Lei 9.656, de 3 de junho de 1998, constitui as Operadoras de Saúde Suplementar. Agência Nacional de Saúde (ANS) foi criada no ano 2000, através da Lei 9.961, de 28 de janeiro de 2000. Com o objetivo de regular o setor de saúde no Brasil, abrangendo a assistência de saúde, enquanto as seguradoras ficariam sob o espectro da Superintendência de Seguros Privados (SUSEP).

De acordo com Filho (1999) o sistema público de saúde brasileiro é muito precário, o que fez crescer a demanda pelo sistema suplementar em mais de 38% em menos de uma década.

Para Oliveira (2014) a deterioração do setor público de saúde e a crescente demanda pelo seu setor privado, também evidenciou os gaps sociais, em que Médice (1992) define como pirâmide da saúde no Brasil, mostrando que a base desta é formada por mais de 85% da população brasileira que não tem acesso aos planos de saúde particulares no país. Atualmente aproximadamente 25,4% da população tem plano de saúde. E o setor vem passando por sua maior queda nos últimos dois anos, com redução em sua taxa de crescimento, representando -2,2% em 2015 e -2,8% em 2016, ocasionando uma perda de mais de 2 milhões de pessoas que cancelaram seus planos. Assim, pode associar isso ao atual momento econômico do país (ANS, 2016).

As Operadoras de Planos de Saúde (OPS) passaram a identificar diversas oportunidades na década de 1990, com variadas precificações que otimizaram suas receitas e lhes protegeram do prejuízo por um determinado período.

Para o Ministério da Saúde (2006) a partir do final da década de 1990 passou-se a haver uma série de batalhas entre os beneficiários dos planos de saúde e as OPS. O argumento desses é de que as Operadoras de Planos de Saúde não os protegem contra imprevistos como perda de renda, necessidade de assistências com valor acima da cobertura oferecida pelas OPS. Já o argumento destas é de que o fato de operarem em um mercado altamente competitivo e expostos a sinistros a todo instante, fazem com que a qualidade dos serviços prestados dependa do equilíbrio financeiro que é estimado por diversos modelos estatísticos, objetivando a maximização de receita e evitando perdas econômico-financeiras.

A construção de um setor da saúde suplementar, cujo principal interesse é a produção de saúde, que seja centrado no usuário, que realize ações de promoção de saúde e prevenção de doenças, que observe os princípios de qualidade, integralidade e resolutividade, que tenha uma concepção inclusivo de todos os profissionais de saúde, que respeite o controle social, que esteja completamente articulado com o Ministério da Saúde e cujo o órgão regulador seja também preocupado com a qualificação de seu processo regulatório (PIETROBON, 2008).

Além disso outro fator importante como apontado por Marujo e Martins (2006) é o envelhecimento da população, o que acarreta em um custo elevado para as Operadoras de Planos de Saúde, que buscam minimizá-los com modelos preditivos de expectativa de vida para precificar suas carteiras da maneira mais precisa possível.

Neste enfoque “a análise do risco de insustentabilidade financeira dos beneficiários da carteira é primordial para a sobrevivência da OPS por ser uma oportunidade de avaliar o desempenho financeiro da empresa, bem como buscar mecanismos de controle se diferenciação perante seus concorrentes, tornando-se assim mais competitiva” (OLIVEIRA, 2014, p.22).

Dentro do contexto de aumentar a competitividade, as OPS barram na Resolução Normativa (RN) 195, que segmenta os planos de saúde em individual ou familiar, coletivo adesão ou empresarial e que não permite a livre negociação entre beneficiários e OPS para a precificação de seus planos.

Dado os argumentos apresentados, este trabalho tem como ponto de partida a seguinte pergunta: Qual a melhor técnica para análise de solvência de carteira de um plano de saúde individual?

Como objetivo geral, esta pesquisa busca identificar as melhores ferramentas estatísticas para análise de solvência. Como objetivos específicos, têm-se: a) analisar o perfil de uma carteira individual de uma OPS; e b) verificar as variáveis que mais afetam a solvência de uma carteira individual.

Para atingir esse objetivo, foi construída uma amostra com 3.423 usuários, de uma contratação de um plano individual e com atendimento em Fortaleza, Manaus e Belém. Empregou-se as técnicas de Regressão Logística, Análise Discriminante, Árvore de Classificação e Regressão e *Random Forest* para que fosse possível analisar a solvência financeira da carteira.

2 REFERENCIAL TEÓRICO

2.1 Mercado brasileiro de saúde suplementar

A saúde suplementar iniciou-se no Brasil na década de 1950, no ABC paulista quando alguns médicos se uniram para dar assistência aos funcionários de determinada empresa, através de pagamentos adiantados por parte destes (OLIVEIRA, 2014).

Mas o ponto de inflexão ocorreu em 1964 quando o Instituto de Aposentadorias e Pensões (IAP) foi desobrigado a dar assistência médico-hospitalar de uma grande indústria automobilística, em troca da desoneração previdenciária da mesma. Assim, com um público desassistido nasceu o primeiro convênio entre cooperativa médica e empresa. Isso foi copiado pelas multinacionais nas décadas seguintes e que permeia até hoje (OLIVEIRA, 2014).

No final da década de 70, a assistência à saúde nas empresas era uma forma de complementar a saúde pública do país. Neste período e no início dos anos 80, os planos de saúde consolidaram sua presença no mercado como uma alternativa de assistência à saúde (PEREIRA, 2015).

Os contratos de saúde não tinham a cobertura mínima regulamentada. Algumas operadoras negavam a realização de procedimentos médico-hospitalares, limitavam o número de consultas, dias de internações e/ou exames. O contrato era firmado entre Operadora e usuários.

A Constituição Federal de 1988, definiu que a saúde é dever do Estado, mas ao mesmo tempo não compete ao mesmo a prestação do saúde. E a Lei de 8.080/90 foi criado o Sistema Único de Saúde (SUS), com atendimento ambulatorial até transplante de órgãos, assim a saúde passa ser direito assegurado pelo Estado.

As OPS são reguladas pela lei nº 9.656/98 e pela medida provisória 2.177-44 que regula a entrada, o funcionamento e a saída das empresas do setor de saúde suplementar, além de transferir para o Poder Executivo a responsabilidade de fiscalização econômica deste ramo. ANS é vinculada ao Ministério Público e atua como órgão independente. Após a regulamentação, das 2.100 Operadoras registradas na ANS no início dos anos 2000, apenas 60% ainda encontravam-se em atividade em 2006. Muitas fecharam as portas ou foram incorporadas por outras (PORTELA, 2010).

E o número de Operadoras está diminuindo ano após ano, conforme o Tabela 1:

Tabela 1 – Quantidade de operadoras no Brasil entre 2007 e 2017

Período	Operadoras		Total
	Médico-hospitalares	Odontológicas	
dezembro, 2007	1.168	408	1.576
dezembro, 2008	1.118	403	1.521
dezembro, 2009	1.088	391	1.479
dezembro, 2010	1.045	366	1.411
dezembro, 2011	1.006	365	1.371
dezembro, 2012	961	359	1.320
dezembro, 2013	915	341	1.256
dezembro, 2014	875	342	1.217
dezembro, 2015	824	326	1.150
dezembro, 2016	789	306	1.095
março, 2017	780	296	1.076

Fonte: ANS 03/2017.

É importante ressaltar também que a venda de planos de saúde é condicionada ao registro destes na ANS, o mesmo é concedido para as operadoras atendam a legislação específica. Além de que as OPS estão impedidas de abrir concordata ou pedir falência. Após a regulamentação do setor privado de saúde, as OPS tentam se adequar às normativas da ANS, inclusive no que diz respeito ao equilíbrio financeiro, uma vez que a ANS aumenta o rol de procedimentos a cada dois anos, proporcionando melhor qualidade e disponibilidade ao usuário, mas por sua vez, aumentando os custos assistenciais das Operadoras.

A RN 195 (2009), segrega os tipos de contratos de saúde em 3 ramificações, Individual ou Familiar, Coletivo Empresarial e Coletivo por Adesão. Os mesmos são comercializados em contratos individuais ou familiares e em contratos coletivos, sendo este último oferecidos a empresas ou associações. Os contratos individuais ou familiares são celebrados com pessoas físicas e suas principais características são os seguintes: os reajustes não são livremente negociados, seu limite superior é estabelecido pela ANS e o contrato não pode ser rescindido unilateralmente pela OPS, exceto nos casos de inadimplência. Os contratos coletivos são celebrados com pessoas jurídicas, seu reajuste é resultados de livre negociação e podem ser rescindidos a qualquer instante, por qualquer uma das partes, desde que observadas as condições estabelecidas no contrato (ALMEIDA, 2009).

A RN 171(2008) determina que o reajuste dos planos Individuais ou Familiares será estabelecido pela média de reajuste dos planos coletivos e divulgado pela ANS. O percentual de reajuste dos planos coletivos é de livre negociação entre Operadoras e Pessoa Jurídica contratante, Nos casos de contratos com quantitativo superior a 30 vidas. Para os contratos com até 30 vidas é aplicado um único percentual, definido pela Operadora,

conforme RN 309, ficando a critério da Operadora incluir empresas com número maior de vidas no Pool de risco. A Lei 9.656/98, proíbe o cancelamento dos contratos individuais, não fazendo a mesma restrição para os coletivos, esses é um dos motivos da Agência não intervir na negociação de reajuste dos planos coletivos.

A legislação é mais intensa nos planos Individuais ou familiar, pois nos planos coletivos, as pessoas jurídicas têm maior poder de barganha. Caso, a ANS continue regulando apenas os planos individuais, pode-se perder o sentido da Agência e há necessidade de ampliação da regulamentação consolidando ações regulatórias na interface público-privado (REIS, 2007).

Atualmente poucas OPS ofertam planos com tipo de contratação Individual ou familiar, uma vez que os planos coletivos são de livre negociação, o que aumenta a lucratividade..

O reajuste estabelecido pela ANS, para os planos individuais são calculados com a mesma metodologia desde 2001, onde é levando em consideração a média dos reajustes aplicados dos planos coletivos superior a 30 vidas, do ano anterior.

A Agência recebe as informações dos reajustes aplicados nos contratos coletivos através de um sistema de envio de informações para ANS denominado Reajuste dos Planos Coletivos (RPC), determinado através da Instrução Normativa N° 13 de 21 de julho de 2006. Deve ser enviado trimestralmente, considerando os reajustes aplicados dentro do trimestre.

Os reajustes dos planos coletivos são reajustados de acordo com as cláusulas contratuais de cada contrato, o que pode ser inclusive considerando o reajuste por sinistralidade, em caso de desequilíbrio financeiro, acrescido da inflação médica.

A Sinistralidade, índice que representa quanto da receita assistencial é consumida pelo custo assistencial. E é calculado, conforme abaixo:

$$\text{Sinistralidade}(\%) = \frac{\text{Custo Assistencial}}{\text{Receita Assistencial}}$$

Júnior (2009) explica algumas metodologias de reajuste por sinistralidade para ser calculado, desconsiderando premissas estatísticas, onde será medido pela variação entre sinistralidade observada e sinistralidade meta.

$$\text{Reaj. por sinist.} = \frac{\text{Sinistralidade Observada}}{\text{Sinistralidade Meta}} - 1$$

Ou podemos calcular considerando margens de segurança estatística.

$$Reaj. \text{ por sinist.} = \frac{S + z_{\gamma} x \frac{\sigma_s}{\sqrt{n}}}{Sinist. Meta} - 1$$

S = Sinistralidade média do período

$z_{\gamma} \sim N(0,1)$ = para um determinado nível de confiança γ

σ_s = desvio padrão da Sinistralidade mês a mês

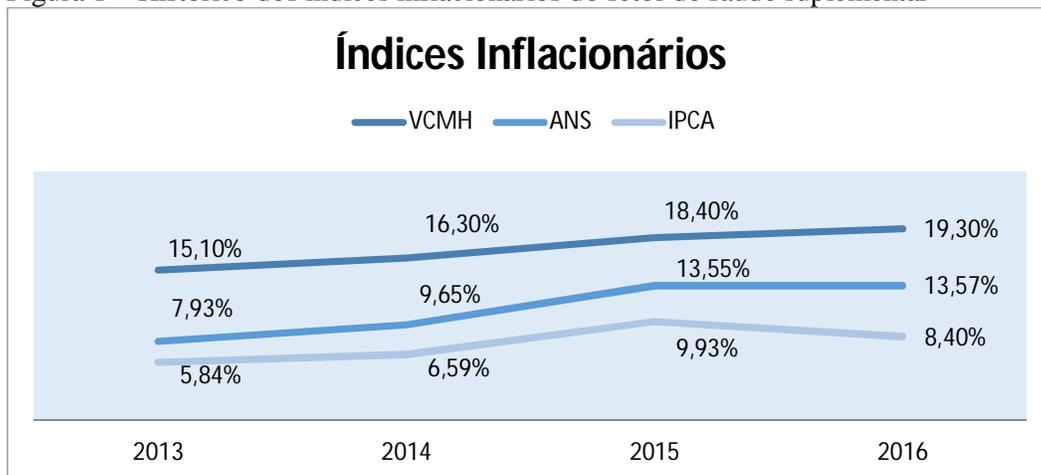
n = número de meses observados

E por fim, o reajuste que será calculado, somente o reajuste por sinistralidade e inflação médica:

$$R = (1 + Reaj. \text{ por sinist.}) \times (1 + \text{inflação médica}) - 1$$

A inflação médica deve ser uma premissa definida em contrato. Cada OPS, pode usar a sua. Porém a inflação médica do setor é medido pelo Variação de Custos Médica Hospitalar (VCMH), divulgada pelo Instituto de Estudos de Saúde Suplementar (IESS). Os reajustes dos planos individuais são historicamente abaixo da inflação do setor, como mostra a Figura 1.

Figura 1 – Histórico dos índices inflacionários do setor de saúde suplementar



Fonte: IESS, ANS e IBGE

O índice de Variação do Custo Médico Hospitalar do IESS – VCMH – expressa a variação do custo médico hospitalar per capita das operadoras de planos de saúde entre dois períodos consecutivos de 12 meses cada (IESS, 2013).

A VCMH é uma medida da variação do custo médico-hospitalar de operadoras de planos e seguros de saúde. O cálculo é feito para um conjunto de planos individuais (antigos e novos) de operadoras que representam cerca de um quarto do mercado (IESS, 2016).

A variação do custo médico-hospitalar (VCMH) é calculada considerando-se o custo médio por beneficiário em um período de 12 meses (média móvel) em relação às despesas médias dos doze meses imediatamente anteriores (IESS, 2016).

Em 2016 o VCMH fechou em 16%, nesse período de análise. O principal componente da VCMH é o grupo de internações, responsável por 60,0% do custo, seguido pelos grupos de Exames Complementares (15,0%), Consultas (11,0%), Outros Serviços Ambulatoriais (OSA) (4,0%) e Terapias (10,0%) (IESS, 2016). Segundo a Associação Brasileira de Planos de Saúdes (Abramge), desde 2007 a saúde suplementar opera com margens operacionais inferiores a 1%", o que resulta na "dificuldade das operadoras em alcançarem o equilíbrio econômico-financeiro necessário.

De acordo com a Federação Nacional de Saúde Suplementar (FenaSaúde), o índice de variação das despesas da assistência à saúde é que determina o reajuste anual das mensalidades dos planos. Esse índice é formado pela combinação de dois componentes básicos: a variação dos preços pagos por consultas médicas, exames complementares, atendimento ambulatorial, terapias e internações; e o aumento da quantidade de procedimentos utilizados por beneficiário desses serviços.

As OPS tentam equilibrar as Despesas Administrativas e Assistenciais, buscando deixar a carteira solvente e com valores competitivo no mercado.

A capacidade de prever insolvência tem papel fundamental. Do ponto de vista econômico, diz-se que uma companhia está insolvente quando o total de seu passivo é superior ao seu ativo, ou seja, quando ela não pode pagar todas as dívidas assumidas mesmo com a liquidez total de seus bens e direitos. Por isso, estabelecer com antecedência quais empresas apresentam propensão à insolvência e quais são financeiramente saudáveis, é de vital importância para credores e para investidores (SOARES E REBOUÇAS, 2014).

Solvência para qualquer empresa significa um nível aceitável de probabilidade de que esta honrará seus compromissos futuros (FERRIS, 2003).

Martins (1994) cita 3 principais objetivos que devem estar relacionados à solvência: continuidade, lucratividade e crescimento. O primeiro está relacionado à vontade da empresa em saldar seus compromissos e continuar a existir. O segundo e o terceiro são fortemente relacionados às empresas que precisam operar com

lucratividade de forma a financiarem seu crescimento, sem que sejam necessários aportes dos acionistas.

A partir da criação de indicadores para solvência das operadoras de planos de saúde, diversos estudos de análise estatística de insolvência começaram a aparecer no meio acadêmico.

Campello (2005), por exemplo, buscou classificar a solvência das OPS de acordo com uma análise multicritério. De acordo com o autor este tipo de análise ajuda a tomar decisões em um cenário com múltiplas variáveis, critérios e características qualitativas.

Campello (2005) conclui que o ELECTRE TRI permitiu-lhe identificar as incomparabilidades ou inconsistências não detectáveis pelos métodos que se baseiam em médias ponderadas. Desta forma verificou que há um comportamento heterogêneo entre as OPS, no cerce de solvência, mesmo que estas busquem o mesmo objetivo.

Já Onuscic *et al* (2007) utilizou uma análise envoltória de dados, verificando-se aqui a evolução das técnicas disponíveis para análise de insolvência, bem como o contínuo estudo nessa área. Os autores desenvolveram um modelo de previsão que melhor classificassem as OPS de acordo com sua solvência.

Onuscic *et al* (2007) selecionaram dez empresas que sofreram processos de concordata entre 1995 e 2003 e para cada uma dessas dez empresas foram sorteadas cinco empresas aleatoriamente que estivessem entre as melhores no setor para isso os autores utilizaram indicadores contábeis e outros por eles construídos. Como resultados chegaram a um *accuracy* de 90% no nível de classificação das empresas quanto à solvência.

2.2 Estudos sobre previsão de insustentabilidade de beneficiários de planos de saúde

Feldman, Wholey e Christianson (1996) analisaram todas as *health maintenance organizations* (HMO) que operaram entre 1986 e 1993. Observaram que durante o período analisado 80 instituições deixaram de existir por conta de fusões, enquanto 149 faliram. Desenvolveram então, por meio de regressão logística multinomial, um modelo para predizer quando uma HMO passaria por um processo de fusão e sobreviveria, ou desapareceria, ou ainda quando viria a falência. Evidenciaram que a quantidade de novos clientes e a rentabilidade são significativas para explicar a fusão ou falência de HMOs.

Fantazzini e Figini (2009) compararam os resultados obtidos por meio de *random forest* e regressão logística aplicados à mensuração do risco de crédito de pequenas e médias empresas. Utilizaram uma amostra composta por 1003 firmas, para as quais foram construídos

16 indicadores financeiros. As duas técnicas foram comparadas por meio de uma análise de curvas ROC. Na etapa de classificação, a capacidade preditiva do *random forest* foi superior ao da regressão logística, obtendo uma medida de 0,9317, contra 0,8513 da segunda técnica. No entanto, na etapa de validação a regressão logística mostrou-se superior, obtendo capacidade preditiva de 0,8414, contra 0,7669 do *random forest*.

Gepp, Kumar e Bhattacharya (2009) utilizaram árvores de classificação e regressão para prever a entrada em falência de empresas. Compuseram uma amostra com 200 empresas, 142 bem-sucedidas e 58 falidas, durante o período de 1971 a 1981. Para a análise foram utilizados 20 indicadores financeiros. Os resultados mostraram que as CART, apesar de serem uma técnica bastante simples, são eficazes para minimizar os erros de classificação em amostras com proporções diferentes de empresas dos dois tipos.

Guimarães e Alves (2009) construíram uma amostra com 597 operadoras que se tornaram insolventes em 2005, isto é, passaram a ter patrimônio líquido negativo. Utilizaram para seu estudo 17 indicadores contábeis calculados a partir de demonstrações financeiras de 2004. Para validação, utilizaram uma amostra com informações referentes ao ano de 2005 de 745 operadoras, 20 das quais se tornaram insolventes em 2006. Obtiveram como resultado, um percentual de acerto de 76,42% para operadoras solventes e 75,00% para insolventes, com precisão geral de 75,71%. Calcularam ainda uma medida de acurácia buscando medir a capacidade preditiva da técnica independente do ponto de corte, obtendo um valor de 68,65%. Compararam ainda os resultados obtidos pelo modelo desenvolvido com o modelo Z de Altman (1968) aplicado à mesma amostra, o qual resultou nos mesmos percentuais de acerto, mas com uma capacidade preditiva independente do ponto de corte ligeiramente menor (68,77%).

Para estudar a previsão de insolvência em operadoras de planos de saúde, Viana *et al.* (2012) construíram uma amostra com 769 operadoras, sendo 15 insolventes e 754 solventes. As operadoras consideradas insolventes foram aquelas que passaram a ter patrimônio líquido negativo em 2009. Utilizaram 25 indicadores contábeis calculados a partir de demonstrativos financeiros de 2008. As variáveis significantes para a classificação foram a relação entre o capital de terceiros e o capital próprio, índice de endividamento, ambos classificados pelos autores como indicadores de capital, e prazo médio de contraprestações a receber, um indicador de ciclo financeiro.

Barros (2013), ao investigar a capacidade preditiva de técnicas de classificação aplicadas à insolvência de operadoras de planos de saúde, considerou quatro amostras, para os

anos de 2009 a 2012, com 782, 752, 721 e 763 operadoras em cada ano, respectivamente. Calculou para isso seis indicadores, um de rentabilidade (Rentabilidade sobre o ativo), um de capital (Endividamento), um de ciclo financeiro (Giro de ativo total), dois de liquidez (Liquidez geral e corrente), e um operacional (Indicador combinado). Comparou o desempenho de seis modelos de classificação existentes na literatura brasileira (ELIZABETSKY, 1976; MATIAS, 1978; KANITZ, 1978; ALTMAN *et al.*, 1979; SANVICENTE E MINARDI, 1998; GUIMARÃES; ALVES, 2009). Segundo a autora, o modelo de Elizabetsky (1976) foi o melhor para prever a insolvência, já que apresentou precisão média de 98,90.

Soares e Rebouças (2014) estudaram técnicas de classificação aplicadas à previsão de insolvência de empresas brasileiras de capital aberto, através de informações contábeis com amostra de 87 companhias. Utilizou as técnicas Análise Discriminante Linear, Regressão Logística, Árvores de Classificação e Redes Neurais Artificiais e a melhor técnica utilizada foi a de Redes Neurais Artificiais.

Oliveira *et al.* (2016) estudou a sustentabilidade financeira dos beneficiários da carteira de uma operadora de planos de saúde cearense composta por 38.875 usuários, ativos, entre os anos de 2011 e 2013. Utilizou para isso as técnicas de regressão logística, as árvores de classificação e o classificador de vizinhos mais próximos. Para comparar essas técnicas, empregou a análise de curvas ROC (*Receiver Operating Characteristic*). Os resultados mostraram que a técnica com melhor capacidade preditiva foi a regressão logística, com área abaixo da curva de 0,72. As variáveis apontadas como mais importantes pelo trabalho foram a Idade e o Tipo de Plano, Receita, Consulta e Odontologia.

Soares (2017) estudou o efeito da idade, do estado de sobrevivência e da proximidade à morte sobre os custos assistenciais em uma Operadora de Plano de Saúde do estado do Ceará e foram utilizadas informações de cerca de 300.000 beneficiários. Foi utilizada a técnica de Árvore de classificação e regressão, *random forest*, teste T e Wilcoxon. O resultado obtido mostra a importância de que o estado de sobrevivência e a proximidade à morte sejam levados em consideração para projeção dos custos assistências em OPS.

3 METODOLOGIA

Neste capítulo são descritos os procedimentos metodológicos utilizados para a realização da pesquisa. Inicialmente, descrevem-se as características da população e as etapas para a obtenção da amostra estudada. Posteriormente, são explicadas as técnicas estatísticas de classificação empregadas no estudo.

3.1 Tipologia da Pesquisa

A pesquisa é de natureza quantitativa. De acordo com Silva e Menezes (2001), é de cunho explicativo, visto que os dados utilizados receberam tratamento estatístico que possibilitaram verificar os fatores que afetam a solvência de uma operadora de saúde.

Quanto aos objetivos, a pesquisa é de natureza descritiva, pois visa descrever as características de determinada população, fenômeno ou o estabelecimento de relações entre as variáveis. (SILVA E MENEZES, 2001). A pesquisa também é de natureza explicativa, pois “visa identificar os fatores que determinam ou contribuem para a ocorrência dos fenômenos” (SILVA E MENEZES, 2001). Já em relação aos procedimentos, este trabalho realizará uma pesquisa bibliográfica, visto que será baseado em estudos já publicados em livros, revistas, periódicos e na internet.

3.2 População e amostra

A Operadora de Saúde, utilizada nesse estudo, tem atualmente 3.891.168 de beneficiários. O período de coleta foi o ano de 2015 e 2016, sendo considerados na amostra usuários com atendimento nas localidades de Fortaleza/CE, Belém/PA e Manaus/AM com contratação dos planos coletivos individual/familiar.

Da carteira selecionada aleatoriamente, foram selecionados os indivíduos que permaneceram no plano durante todo o período de análise, totalizando 3.243 observações. Para a execução da análise proposta, a amostra total foi subdividida, por meio de amostragem probabilística em duas subamostras, uma com 1.622 beneficiários, utilizada para o treinamento dos modelos estatísticos propostos, e outra com 1.621, que serviu para a validação dos resultados obtidos.

Estratégia de análise

O índice de sinistralidade de 2016 é variável dependente. Se a sinistralidade do usuário for acima de 75% é considerado beneficiário insustentável e se a sinistralidade for abaixo de 75% é considerado beneficiário sustentável.

Sexo, idade, custo, receita e cidade de 2015 são variáveis independentes. A operacionalização das variáveis podem ser vista no Quadro 1:

Quadro 1 – Variáveis utilizadas no estudo

Variável	Tipo	Descrição	Categorias
Sustentabilidade	Qualitativa	Relação entre o Custo e a Receita anuais dos beneficiários. São considerados sustentáveis aqueles com índice inferior à 0,75.	Duas categorias (0 - Sustentável e 1 - Insustentável)
Idade	Quantitativa	Idade, em anos do beneficiário	
Sexo	Qualitativa	Sexo do beneficiário)	Duas categorias (0 - Feminino e 1 - Masculino)
Cidade	Qualitativa	Cidade de residência do beneficiário	Duas variáveis <i>dummy</i> para representar os beneficiários de Fortaleza, Belém e Manaus
Receita	Quantitativa	Valor pago pelo beneficiário em um ano	
Custo	Quantitativa	Custo gerado pelo beneficiário em um ano	

Fonte: Elaborada pela autora.

As variáveis foram analisadas por meio de estatística descritiva e inferencial, para uma noção inicial do seu comportamento. Posteriormente, foram aplicadas técnicas de classificação.

3.3.1 Regressão Logística

A Regressão Logística é uma técnica utilizada em modelos estatísticos e que busca melhor adequação das variáveis ao modelo e maximiza o princípio da parcimônia. A grosso modo a regressão logística é muito similar à regressão linear, em que há um modelo que busca explicar as relações entre uma variável dependente (ou variável resposta) e variáveis independentes (ou explanatórias). Mas a principal diferença entre esses dois tipos de análises é que na regressão linear, a variável declarada dependente é contínua, enquanto nos modelos de regressão logística esta variável é binária ou dicotômica (HOSMER;

LEMESHOW, 2000).

Algumas considerações teóricas e práticas sugerem que, quando a variável resposta é binária, a forma da função resposta será frequentemente curvilínea. Para esses casos a função logística é a mais utilizada. Sua expressão matemática é dada por:

$$E[Y] = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} = [1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_p X_p)]^{-1}$$

em que Y é a variável resposta, X_1, \dots, X_p são as variáveis explicativas e β_0, \dots, β_p são os coeficientes estimados a partir do conjunto de dados por meio da máxima verossimilhança.

Uma propriedade bastante útil da função logística é o fato dela poder ser linearizada. Definindo $\pi(x) = E[Y|x] = P(Y=1|x)$, pode ser aplicada a transformação *logit*, dada por:

$$\text{logit}[\pi(x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right)$$

logo, a função resposta *logit* pode ser expresso como:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

A razão entre $\pi(x)$ e $1 - \pi(x)$ é chamada de *odds ratio*, ou razão de chance, pois representa uma relação entre as probabilidades de sucesso e fracasso.

De acordo com Corrar *et al* (2012), a técnica foi desenvolvida durante a década de 1960, sendo o *Framingham Heart Study*, realizado pelo National Heart, Lung and Blood Institute – NHLBI e a Universidade de Boston, um dos primeiros estudos que mais contribuíram para sua notoriedade. A RL foi utilizada para identificar fatores responsáveis pelo desencadeamento de doenças cardiovasculares.

3.3.2 Análise Discriminante

A Análise Discriminante é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes (KHATTREE & NAIK, 2000). A discriminação ou separação é a primeira etapa, sendo a parte exploratória da análise e consiste em se procurar características capazes de serem utilizadas para alocar objetos em diferentes grupos previamente definidos (VARELLA, 2013).

Trata-se de uma técnica estatística multivariada para a classificação de objetos, bastante utilizada quando a variável dependente do modelo é qualitativa e as independentes são contínuas.

O objetivo desta análise é identificar as variáveis que melhor discriminam a

variável dependente. A primeira etapa deste modelo é a discriminação das variáveis a partir da equação abaixo:

$$\tau = \frac{SQ_{dg}}{SQT}$$

Em que : τ é a razão entre a soma dos erros ao quadrado dentro de cada variável (SQ) e a soma dos erros ao quadrado total SQT, variando de 0 a 1 e que quanto mais próximo de 1 indica ausência de diferença entre as variáveis. (CUNHA, 2014).

Após esta etapa discriminatória, o próximo passo da análise discriminatória é identificar as funções discriminantes, dada por meio de uma regressão linear:

$$Z_n = \beta_0 + \beta_i X_i$$

Onde:

Z é a variável dependente;

β_0 é o intercepto do modelo;

β_i são os coeficientes discriminantes de cada variável i e X é a matriz de variável explicativa.

O último passo da análise discriminatória é aplicar as funções discriminantes nas matrizes de soma dos erros quadrados e soma dos erros ao quadrado totais, buscando a maximização de resposta do modelo a partir dos scores de corte gerados pelas funções discriminantes.

Assim a análise discriminante será utilizada neste trabalho para escolher as variáveis que melhor se encaixam no modelo de insustentabilidade financeira.

3.3.3 Árvore de Classificação e Regressão

O algoritmo de Árvore de Classificação e Regressão (*Classification and Regression Trees – CART*), foi proposto por Breiman *et al* (1984) e consiste em uma técnica não paramétrica que induz tanto árvores de classificação, caso a variável dependente seja categórica, quanto para árvores de regressão, sendo a variável dependente contínua. Ainda segundo o autor, uma das maiores virtudes da CART é a capacidade de pesquisa de relações entre os dados, mesmo que não sejam evidentes.

O método CART baseia-se na execução de partições binárias sucessivas de uma amostra, com base nos resultados amostrados das covariáveis, buscando a constituição de subamostras internamente homogêneas. A classificação dessas subamostras é realizada

conforme alguma medida descritiva e a predição de novos elementos, executada por meio da estrutura de classificação constituída (TACONELI; ZOCCHI; DIAS, 2009).

Os componentes elementares do modelo são os nós e as regras de divisão (*splitting rules*). O primeiro nó de uma árvore é chamado de raiz e representa todo o conjunto de dados. Os nós terminais recebem o nome de folhas. Os nós que dão origem a outros são chamados de pais, enquanto aqueles gerados são denominados de filhos.

Timofeev (2004) aponta algumas vantagens das CART. Quando o método é utilizado para classificação, o próprio algoritmo indica quais variáveis são mais importantes, descartando aquelas menos significantes, o que é bastante útil quando não se tem conhecimento prévio de quais variáveis são mais relevantes para a classificação a ser realizada.

As CART também lidam com *outliers* de maneira muito mais robusta do que outros métodos, já que o método permite que esses tipos de observações sejam isolados em um nó a parte das demais observações. Esta é uma característica muito importante já que *outliers* costumam ter efeitos negativos sobre os resultados de modelos estatísticos.

3.3.4 *Random Forest*

O *random forest* é um algoritmo proposto por Breiman em 2001, o qual é capaz de gerar, de maneira aleatória, diferentes árvores de classificação em diferentes subespaços amostrais obtidos por meio do método de amostragem *bootstrap* (NISBET; ELDER; MINER, 2009). A classificação baseia-se pela moda das árvores de decisão geradas pelo algoritmo.

Segundo Dean (2014), o *random forest* é um método que reúne um grande número de árvores, consideradas classificadores mais fracos, e as usa para construir, de modo coletivo, um classificador que seja melhor e mais consistente, a floresta. Segundo o autor, isso se relaciona com o axioma que diz que o todo é maior que a soma das partes, na ideia de que a classificação por *random forest* é superior à utilização de várias árvores de classificação separadamente.

3.3.5 *Curva ROC*

Como o estudo se propõe a avaliar o desempenho do modelo, uma estratégia de comparação entre os métodos utilizados se faz necessária. É comum que se utilize as precisões gerais obtidas para cada modelo construído, como medida de desempenho dos

mesmos. Esse tipo de análise, no entanto, não é indicado para os casos em que existam desproporções entre as classes que compõe a amostra, situação evidenciada nesse trabalho.

Para fazer esse tipo de análise é possível utilizar curvas de ROC, do inglês *Receiver Operating Characteristic*. Esse tipo de análise baseia-se na relação entre a sensibilidade e a especificidade de classificadores binários. A sensibilidade é a taxa de verdadeiros positivos (*true positive rate*), que nesse estudo é a probabilidade de beneficiário insustentável ser classificado como tal. O segundo atributo, a especificidade, é a taxa de verdadeiros negativos (*true negative rate*), a probabilidade de um beneficiário sustentável ser classificado como tal.

De acordo com Flach (2010), se um modelo de classificação um escore que seja proporcional ao grau de certeza com o qual determinada entrada pertença à classe positiva, é possível determinar vários pontos de corte, os quais definirão diferentes proporções para a taxa de verdadeiros positivos e falsos positivos (*false positive rate*), que nesse estudo pode ser entendida como classificar uma empresa insolvente como solvente. Observando todos os possíveis pontos de corte desde zero até um e ligando-os todos, forma-se uma curva composta por segmentos, a qual recebe o nome de Curva de ROC.

4 RESULTADOS

4.1 Análise descritiva e inferencial

Inicialmente, estudou-se como se dá o comportamento dos grupos de beneficiários sustentáveis e insustentáveis em cada uma das variáveis isoladamente. Na Tabela 1 é apresentado um comparativo entre os grupos em relação ao sexo.

Tabela 2 – Distribuição dos beneficiários por estado de insustentabilidade e sexo

		Sexo			Teste quiquadrado
		Feminino	Masculino	Total	
Insustentável	Não	1.309	907	2.216	2,113 (0,146)
	Sim	635	392	1.027	
	Total	1.944	1.299	3.243	

Fonte: Elaborada pela autora.

Percebe-se que há uma concentração maior de beneficiários do sexo feminino (59,94% da amostra, contra 40,06% de usuários do sexo masculino). No entanto, não se pode afirmar que haja uma aderência entre o sexo do beneficiário e sua sustentabilidade. Isso é evidenciado pela proporção semelhante de beneficiários sustentáveis e insustentáveis em cada um dos sexos, uma vez que 67,34% das mulheres e 69,82% são insustentáveis, e corroborado pelo teste quiquadrado, a medida que não é possível rejeitar a hipótese nula de que as duas variáveis são independentes com 10% de significância (valor p superior a 0,10).

Na Tabela 2 é apresentada a distribuição dos beneficiários sustentáveis e insustentáveis de acordo com a cidade onde moravam durante o período de análise.

Tabela 3 – Distribuição dos beneficiários por estado de insustentabilidade e cidade

		Cidade				Teste quiquadrado
		Belém	Fortaleza	Manaus	Total	
Insustentável	Não	957	895	364	2.216	10,681 (0,000)
	Sim	414	474	139	1.027	
	Total	1.371	1.369	503	3.243	

Fonte: Elaborada pela autora.

Há uma concentração similar de beneficiários de Belém (42,28%) e Fortaleza (42,21%), enquanto Manaus tem uma representatividade menor (15,51%). Diferentemente do sexo, o teste quiquadrado realizado para relação entre a cidade do beneficiário e seu estado de sustentabilidade, indica que há uma aderência entre as duas variáveis, uma vez que o teste foi significativo a 1% (valor p de 0,000). Ao analisar os percentuais de beneficiários sustentáveis e insustentáveis por cidade, percebe-se que há, percentualmente, uma maior concentração de beneficiários insustentáveis em Manaus (72,37%), seguida de Belém (69,80%) e Fortaleza (65,38%).

Por fim, foram comparados os grupos de beneficiários sustentáveis e insustentáveis quanto à idade, custo e receita gerados para o plano. Os resultados são apresentados na Tabela 3.

Tabela 4 – Estatísticas descritivas da idade, custo e receita dos beneficiários por grupo

	Sustentável			Insustentável			Teste T
	Média	Mediana	Desvio	Média	Mediana	Desvio	
Idade	21,16	17,00	19,10	15,62	8,00	16,67	8,401 (0,000)
Custo	745,66	229,99	3.049,36	1.511,62	849,59	3.191,76	-6,447 (0,000)
Receita	1.422,03	1.145,17	1.411,36	1.293,92	1.229,96	996,96	2,965 (0,003)

Fonte: Elaborada pela autora.

Para o comportamento da idade, é possível perceber que há uma concentração de usuários mais jovens no grupo de beneficiários insustentáveis, tendo em vista que esse grupo apresentou média de idade de 15,62 e 50% dos integrantes desse grupo têm até 8 anos de idade. Já os sustentáveis têm média de idade de 21,16 e 50% do grupo com 17 anos ou menos.

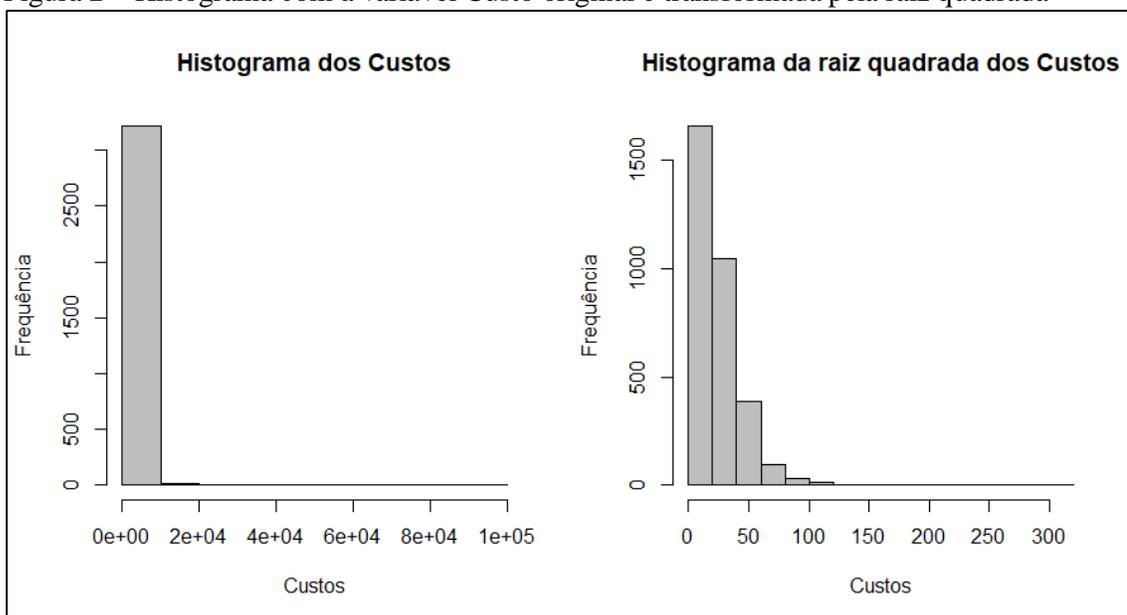
Para o comportamento dos custos, nota-se que os beneficiários sustentáveis geraram em média menos custos para o plano (R\$ 745,66), com 50% desses beneficiários com um custo igual ou inferior a R\$ 229,99. Os beneficiários insustentáveis geraram em média R\$ 1.511,62 com 50% desses usuários com receita menor ou igual a R\$ 849,59.

Nas receitas, o que se pode observar é que os usuários sustentáveis geraram em média maior receita (R\$ 1.422,03), com 50% desse grupo abaixo do patamar de 1.145,17, enquanto o grupo de insustentáveis teve uma média de receita de R\$ 1.293,92, e R\$ 1.229,96 para o percentil 50.

Destaca-se ainda que as diferenças entre os dois grupos para as três variáveis são estatisticamente significantes a 1%, analisando as significâncias obtidas para o teste T. Já para o teste de Wilcoxon, não se pode afirmar que a diferença entre os beneficiários sustentáveis e insustentáveis é estatisticamente significativa.

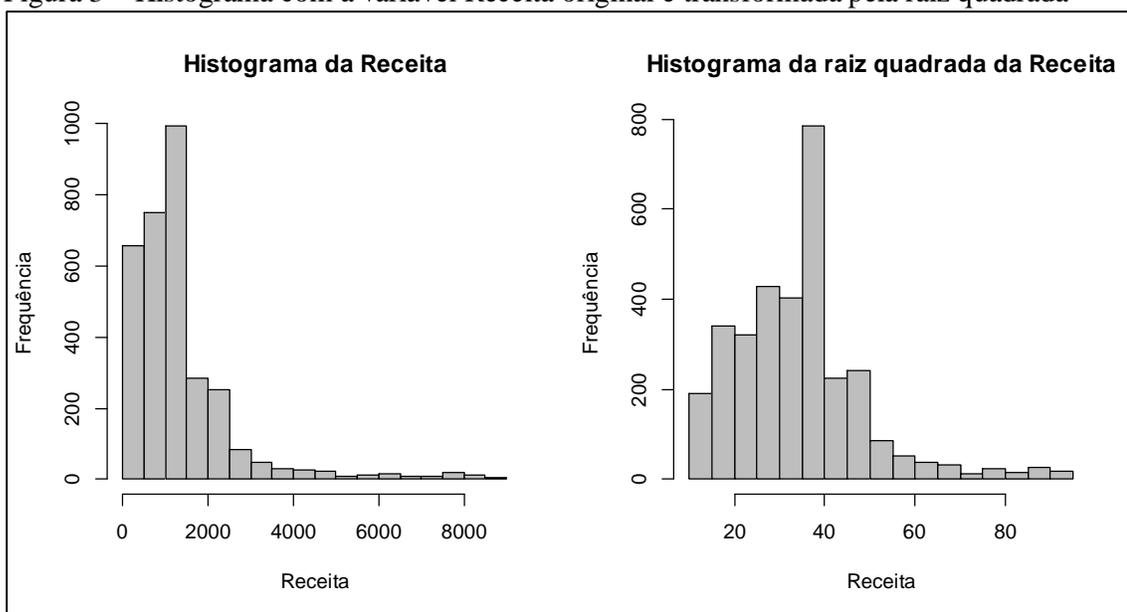
A diferença entre a média e mediana para as variáveis na Tabela 3 indica a existência de *outliers* que podem afetar os resultados das análises multivariadas realizadas. Além disso, a escala das variáveis Receita e Custo (medidas em centenas e milhares) estão em uma escala maior quando comparadas com a variável Idade (medida em dezenas). Por esses motivos, as variáveis foram transformadas por meio da aplicação da raiz quadrada, cujo objetivo é suavizar sua distribuição. O efeito da aplicação dessa transformação pode ser observado por meio dos histogramas na Figura 2 para o custo e na Figura 3 para a receita.

Figura 2 – Histograma com a variável Custo original e transformada pela raiz quadrada



Fonte: Elaborada pela autora.

Figura 3 – Histograma com a variável Receita original e transformada pela raiz quadrada



Fonte: Elaborada pela autora.

4.2 Análise multivariada

Após o estudo preliminar da amostra, foram aplicadas as técnicas estatísticas para a classificação dos beneficiários quanto ao seu estado de sustentabilidade para o plano. A

primeira técnica aplicada, devido à maior frequência de utilização para esse tipo de problema, foi a regressão logística. A classificação dada pelo modelo é apresentada na Tabela 4.

Tabela 5 – Classificação dada pela regressão logística

Regressão Logística		Classificados		Total
		Sustentáveis	Insustentáveis	
Observado	Sustentáveis	752 67,87%	356 32,13%	1108 100,00%
	Insustentáveis	165 32,16%	348 67,84%	513 100,00%
Total		917	704	1621

Fonte: Elaborada pela autora.

A regra de classificação estimada pela regressão logística foi capaz de classificar corretamente 67,87% dos beneficiários sustentáveis na amostra de validação, enquanto 67,84% dos beneficiários insustentáveis foram classificados corretamente. A acurácia da técnica foi de 67,86% considerando os dois grupos.

A seguir, foi empregada a análise discriminante, cuja classificação é apresentada na Tabela 5.

Tabela 6 – Classificação dada pela análise discriminante

Análise Discriminante		Classificados		Total
		Sustentáveis	Insustentáveis	
Observado	Sustentáveis	755 68,14%	353 31,86%	1108 100,00%
	Insustentáveis	165 32,16%	348 67,84%	513 100,00%
Total		920	701	1621

Fonte: Elaborada pela autora.

Nessa técnica, foi obtida uma precisão de 68,14% para os beneficiários sustentáveis e 67,84% para os insustentáveis, com uma precisão geral de 68,04%, ao serem considerados os dois grupos.

A terceira técnica aplicada foi a árvore e classificação e regressão, cujos resultados são apresentados na Tabela 6.

Tabela 7 – Classificação dada pela árvore de classificação e regressão

Árvore de Classificação		Classificados		Total
		Sustentáveis	Insustentáveis	
Observado	Sustentáveis	931 84,03%	177 15,97%	1108 100,00%
	Insustentáveis	299 58,28%	214 41,72%	513 100,00%
Total		1230	391	1621

Fonte: Elaborada pela autora.

Para essa técnica observa-se uma classificação correta de 84,03% dos beneficiários sustentáveis e 41,72% dos beneficiários insustentáveis. A precisão geral dessa técnica atingiu o patamar mais elevado (70,64%). Apesar disso, deve-se atentar para o fato de que a precisão dessa técnica se concentra majoritariamente nos beneficiários sustentáveis (que compõem a maior parte da amostra), enquanto nos insustentáveis a técnica acerta menos de 50% no ponto de corte adotado. Ou seja, apesar do aparente desempenho superior da árvore de classificação, sua incapacidade de detectar os beneficiários insustentáveis, acaba por torná-la menos indicada que as técnicas anteriores.

Por fim, a última técnica empregada foi o *random forest*, cujos resultados são apresentados na Tabela 7.

Tabela 8 – Classificação dada pelo *Random Forest*

<i>Random Forest</i>		Classificados		Total
		Sustentáveis	Insustentáveis	
Observado	Sustentáveis	723 65,25%	385 34,75%	1108 100,00%
	Insustentáveis	179 34,89%	334 65,11%	513 100,00%
Total		902	719	1621

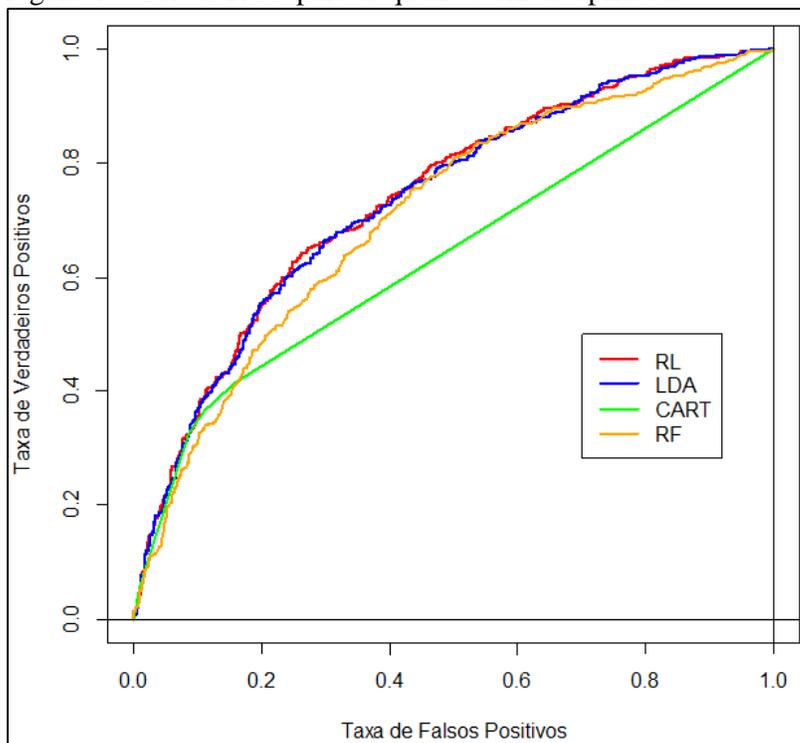
Fonte: Elaborada pela autora.

A técnica de *random forest* classificou corretamente 65,25% dos beneficiários sustentáveis na amostra de validação, e 65,11% dos beneficiários insustentáveis. A precisão geral da técnica foi de 65,21%.

Com o objetivo de avaliar melhor as técnicas de classificação utilizadas, foi empregada a curva ROC, cuja principal medida de avaliação é a área abaixo da curva (AUC). Como a AUC representa a probabilidade de acertar uma classificação aleatória, considera-se essa uma medida mais indicada para avaliar o desempenho desse tipo de técnica. Ao basear a avaliação apenas na precisão, podem surgir problemas como o que foi evidenciado na Tabela 6 para a árvore de classificação. Como a maior parte da amostra é composta por beneficiários sustentáveis, uma técnica de classificação que acerte mais nesse grupo, terá o seu índice de precisão geral inflado. A AUC então sana esse tipo de problema.

As curvas geradas para as quatro técnicas são apresentadas na Figura 4, e os resultados para a AUC estão na Tabela 8.

Figura 4 – Curvas ROC para as quatro técnicas aplicadas



Fonte: Elaborada pela autora.

Tabela 9 – Área abaixo da curva para as quatro técnicas empregadas

Técnica	Área Abaixo da Curva (AUC)
Regressão Logística (RL)	0,73882
Análise Discriminante (LDA)	0,73465
Árvore de Classificação (CART)	0,63827
<i>Random Forest</i> (RF)	0,70988

Fonte: Elaborada pela autora.

Com base nos resultados, é possível perceber que a técnica com melhor capacidade preditiva é a Regressão Logística, com 73,88% de chance de acertar a classificação de uma observação aleatória, seguida da Análise Discriminante, com probabilidade de acerto de 73,47%. A Árvore de Classificação, apesar de ter apresentado a melhor precisão geral, obteve o pior desempenho relativo à AUC, com 63,83% de probabilidade de acertar uma nova classificação, corroborando com o que foi exposto que o fato da técnica concentrar seus acertos no grupo de sustentáveis, acaba por inflar sua precisão.

5 CONCLUSÃO

A saúde suplementar tem grande representatividade, como atividade econômica, em diversos países. No Brasil apenas 24% da população tem acesso a esse sistema de saúde, tendo que recorrer, em grande maioria, ao sistema de saúde universal. Mas a melhora no cenário econômico brasileiro durante a década de 2000 fez a demanda por planos de saúde particular crescer a taxas surpreendentes.

Nesse contexto, o mercado de prestação de serviços de saúde ficou muito atrativo e mais Operadoras de Planos de Saúde nasceram no âmbito nacional. Mas estas OPS tem tido barreiras econômicas muito fortes, impostas pela Agência Nacional de Saúde (ANS), como a segregação dos planos de saúde em categorias como familiar, individual e coletivo. O que inviabiliza a clusterização dos clientes e uma precificação mais personalizada para cada cluster de clientes. Isto acaba afetando a competitividade entre as OPS, dado que as mesmas precisam precificar seus planos de acordo com suas carteiras de clientes, colocando em risco a solvência das mesmas. Que em face a essas barreiras precisam melhorar de forma contínua suas técnicas de análises de solvência.

Para realização do estudo, foi analisado a carteira do plano individual das localidades de Fortaleza, Manaus e Belém, dos anos de 2015 e 2016.

Assim, através de diversas técnicas estatísticas como a regressão logística, análise discriminante, árvore de classificação e o *random forest* foi possível identificar o melhor modelo para análise de solvência. As técnicas com melhor poder preditivo foram a regressão logística e a análise discriminante, com capacidade preditiva de 73,88% e 73,47%, respectivamente, segundo a área abaixo da curva ROC.

Ainda foi possível analisar a carteira individual de uma OPS com atuação no Norte e Nordeste, comparando variáveis sociodemográficas e de gênero. Por fim o trabalho alcançou mais um objetivo que foi o de verificar as variáveis que impactam a solvência de uma carteira. Sendo a idade, o custo e a receita gerada pelo planos de saúde os fatores que mais impactam na solvência de uma OPS.

Os achados neste trabalho contribuem para o aumento na literatura sobre a solvência de OPS, bem como para as práticas utilizadas pelas operadoras no momento de definir suas estratégias de mercado e/ou *target* buscando variar cada vez mais as carteiras balanceando-as com as variáveis que afetam sua solvência.

As limitações encontradas neste trabalho consistem na indisponibilidade de dados de outras OPS, tanto nacional quanto internacional, o que inviabilizou a análise comparativa com relação a regiões do Brasil e a diferença entre as OPS brasileiras e estrangeiras.

Sugere-se, a partir dos fatores expostos acima, que este estudo sirva como base comparativas para a demais técnicas estatísticas que possam sobressair em relação às apresentadas neste trabalho. Esta pesquisa pode ainda servir com base de comparação para solvência de outras carteiras de OPS espalhadas pelo Brasil e pelo mundo.

REFERÊNCIAS BIBLIOGRÁFICAS

_<<http://www.ans.gov.br/aans/quem-somos/historico>> Acesso em: 25/10/2017.

_<<http://www.ans.gov.br/perfil-do-setor/dados-gerais>> Acesso em: 25/10/2017.

_ Análise Multivariada Aplicada as Ciências Agrárias – Análise Discriminante. **Nota de Aula**. 2013. Disponível em

<<http://www.ufrj.br/institutos/it/deng/varella/Downloads/multivariada%20aplicada%20as%20ciencias%20agrarias/Aulas/ANALISE%20DISCRIMINANTE.pdf>>.

Acessado em 10/05/2017.

ALMEIDA, Renata Gasparello de; SANT'ANNA, Annibal Parracho. Composição Probabilística na Avaliação do Risco de Operados de Planos de Assistência à Saúde.

Dissertação. Universidade Federal Fluminense. 2009. Disponível em

<http://www.producao.uff.br/antigo/conteudo/rpep/volume92009/RelPesq_V9_2009_04.pdf>. Acessado em 04/04/2017

CAMPELO, Ricardo Camacho. Análise Multicritério aplicada a classificação da solvência de operadoras de assistência à saúde. **Dissertação**. 2005. Disponível em

<http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Materiais_por_assunto/dissertacoes-analisemulticriterioaplicadaaclassificacaodasolvenciadeoperadorasdeassistenciaasaude.pdf>.

Acessado em 13/03/2017.

CUNHA, Franciely Farias. Análise Discriminante. **Monografia**. 2014. Disponível em

<http://www.ufpa.br/heliton/arquivos/aplicada/seminarios/M1_06_Analise_Discriminante_Franciely.pdf>. Acesso em 20/05/2017.

FILHO, Luiz Tavares Pereira. Iniciativa privada e saúde. Scielo. 1999. Disponível em

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141999000100011>. Acessado em 09/02/2017

FLACH, Peter A, **ROC Analysis: encyclopedia of machine learning**. Nova Iorque: Springer: 2010.

HOSMER, David. W; Lemeshow, Stanley. **Applies logist regression**. 2. Es. New York: Jonh Wiley and Sons, 1989.

HOSMER, David, W; LEMESHOW, Stanley. **Applied Logistic Regression**. John Willey & Sons, INC. 2º edição. 2000

JÚNIOR, José Nazareno Maciel. As Principais Atividades brasileiras dos Atuários nas Operadoras de Saúde. **Nota de Aula**. 2009.

KING, Gary; ZENG, Langche. Logistic Regression in Rare Events Data. Political Analysis. 2001. Disponível em

<<https://dash.harvard.edu/bitstream/handle/1/4125045/relogit%20rare%20events.pdf?s>>. Acesso em 18/05/2017.

MARGOTTO, Paulo R., **Curva ROC Como fazer e interpretar no SPSS**, 2010. Disponível em

MARTINS, Macus V.L. (1994). **Uma Proposta de Metodologia de Avaliação de Solvência de Empresas de Seguro não vida**. RJ – RJ. Dissertação de Mestrado em Administração – COPPEAD – UFRJ.

MARUJO, Ernesto Cordeiro (coord.); MARTINS, Carina Burri (elab.). **Fatores explicativos do aumento de gastos com saúde: breve revisão da literatura internacional**. São Paulo: Instituto Estudos de Saúde Suplementar, 2006.

MÉDICI, A. C. **A medicina de grupo no Brasil**. Brasília: OPAS/Representação do Brasil, 1992. (Série Desenvolvimento de Políticas Públicas, n.1).

NASCIMENTO, Juliana Oliveira, **A importância da ANS – Agência Nacional de Saúde Suplementar, no âmbito da saúde suplementar no Brasil**, 2017. Disponível em: <http://www.ambito-juridico.com.br/site/index.php?n_link=revista_artigos_leitura&artigo_id=3040>. Acesso em: 25/10/2017.

OLIVEIRA, Daniele Adelaide Bradão de. Risco de Insustentabilidade financeira dos beneficiários de uma operadora de plano de saúde: uma comparação de modelos de classificação. **Dissertação**. Universidade Federal do Ceará. 2014

ONUSCIC, Luciana Massaro; NOVA, Silvia Pereira de Castro Casa; ALMEIDA; Fernando Carvalho. Modelos de previsão de insolvência utilizando a análise por envoltória de dados: aplicação a empresas brasileiras. **Revista de Administração Contemporânea**. Vol 11. 2007. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-6552007000600005>. Acessado em: 15/03/2017

PIETROBON, Louise; PRADO, Martha Lenise do; CAETANO; João Carlos. Saúde suplementar no Brasil: o papel da Agência Nacional de Saúde Suplementar na regulação do setor. **Physis: Revista de Saúde Coletiva**. Vol 18. 2008. Disponível em <www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-73312008000400009>. Acessado em: 23/03/2017

PEREIRA, Márcia Angélica Mendes, **Previsão de Insolvência nas Operadoras de plano de saúde no Brasil: Um estudo com técnicas de análise multivariada**, 2015. Monografia

PORTELA, Paulo Roberto, **Saúde Suplementar: o X da questão no Brasil**, 2010. Disponível em <<http://cadernosdeseguro.funenseg.org.br/secao.php?materia=363>>. Acesso em: 25/10/2017.

SOARES, Rômulo Alves, **Efeito da Idade, do Estado de Sobrevivência e da Proximidade à Morte nos Custos Assistenciais de Uma Operadora de Plano de Saúde**, XVII International Conference in Accounting – USP, 2017

SOARES, Rômulo Alves; REBOUÇAS, Sílvia Maria Dias Pedro. Avaliação de Desempenho de Técnicas de Classificação Aplicadas à Previsão de Insolvência de Empresas de Capital Abertos Brasileiras. Revista ADM.MADE. 2014