Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

ANÁLISE DE SINISTRALIDADE DE UMA OPERADORA ODONTOGOLICA NO ESTADO DE MINAS GERAIS

Géssica Silva Marques

Uberlândia-MG
2020

Géssica Silva Marques

ANÁLISE DE SINISTRALIDADE DE UMA OPERADORA ODONTOGOLICA NO ESTADO DE MINAS GERAIS

Trabalho de conclusão de curso de graduação apresentado à Faculdade de Matemática da Universidade Federal de Uberlândia (UFU) como requisito parcial para a obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Lúcio Borges de Araújo

Uberlândia-MG
2020

Universidade Federal de Uberlândia
Faculdade de Matemática

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ___ de _____ de 2020

BANCA EXAMINADORA

_____

Prof. Dr. Lúcio Borges de Araújo

_____

Prof. Dra. Priscila Neves Faria

_____

Prof. Dra. Maria Imaculada de Sousa Silva

Uberlândia-MG
2020

AGRADECIMENTOS

A Deus que me deu saúde e vida para concluir esta jornada.

A minha mãe Maria e minha falecida avó Joana que sempre estiveram comigo e acreditaram que eu teria capacidade de concluir mais esta etapa.

Agradeço ao meu pai Otacílio e meu irmão Jonatas que me deram apoio.

A Universidade Federal de Uberlândia pelo ensino e compromisso tão nobre de repassar conhecimento e pela oportunidade de concluir minha graduação, meus agradecimentos especiais aos professores da Faculdade de Matemática.

Ao meu orientador Professor Lúcio Borges, pelos ensinamentos, orientações neste trabalho e pela paciência. Agradeço pelas aulas ministradas que foram muito importantes durantes os anos de estudo e que me fizeram valorizar ainda mais esta oportunidade.

Agradeço à banca examinadora Professoras Priscila Neves e Maria Imaculada de Sousa Silva pelos ensinos e por prestigiarem minha defesa.

Aos meus colegas da graduação, amigos, familiares, colegas do trabalho que acreditaram em mim e a minha gerência que proporcionou flexibilidade para desempenhar minhas atividades acadêmicas.

RESUMO

Diante da perspectiva de crescimento dos custos na saúde, do envelhecimento populacional e com a constante busca pelos planos de saúde e odontológicos, a Agência Nacional de Saúde criou alguns processos e controles para auxiliar as operadoras na gestão econômica, financeira, operacional e regulatória a fim de manter o equilíbrio entre a oferta e a procura por planos privados. Este trabalho pretende abordar a avaliação do tema sinistralidade, sendo uma relação entre as despesas da operadora e receitas, sendo uma métrica que auxilia nesta gestão financeiro-econômica. Para as análises foram coletados dados de uma operadora odontológica do estado de Minas Gerais, considerando a utilização do plano por beneficiários deste estado e avaliando o período de junho/2019 á julho/2020. Foram incorporadas variáveis *dummys* ao modelo inicial, realizada avaliação por meio do método de regressão linear múltipla utilizando Stepwise, análise de multicolinearidade das variáveis, ajuste dos dados por modelos de Box-Cox e análise de normalidade para definição do modelo que melhor explique a variável resposta sinistralidade, sendo este indicador altamente importante para saúde financeira da operadora. Ao final deste trabalho observa-se que variáveis antes não consideradas para a operadora foram fatores que apresentaram relevância para controle do sinistro.

Palavras-chave: Regressão Linear Múltipla, Stepwise, Multicolinearidade, *Dummys*.

ABSTRACT

From prospect of rising health care costs, an aging population and the constant search for health and dental plans, the National Health Agency created some processes and controls to assist operators in financial, operational and regulatory economic management in order to maintain the balance between supply and demand for private plans. This work intends to approach the evaluation of the accident theme, being a relationship between the operator's expenses and revenues, being a metric that helps in this financial-economic management. For the analysis, data were collected from a dental operator in the state of Minas Gerais, considering the use of the plan by users in this state and an assessment of the period from June / 2019 to July / 2020. Dummy variables were incorporated into the initial model, an evaluation was carried out through the method of multiple linear regression using Stepwise, multicolinearity analysis of the variables, adjustment of the data by Box-Cox models and normality analysis to define the model that best explains the variable response sinistrality, this indicator being highly important for financial health operator. At the end of this work, it is observed that variables previously not analyzed for an operator of external factors that are the source of control of the sinister.

Keywords: Multiple Linear Regression, Stepwise, Multicolinearity, Dummys.

LISTA DE TABELAS

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

Implementada em 1988, a Constituição da República Federativa do Brasil conforme [1] reconhece a saúde como dever do Estado e direito de todos, procura garantir que políticas econômicas e sociais sejam promovidas a fim de obter redução das doenças e acesso igualitário às ações que promovam saúde. Para assegurar que as políticas públicas de saúde sejam implementadas, criou-se o Sistema Único de Saúde (SUS) que possui carácter complementar aos meios privados que oferecem tratamentos médicos e odontológicos, sendo assim incorporados ao setor denominado de Saúde Suplementar.

Apesar de ser um direito constitucional de garantia igualitária à saúde sem distinção de renda, a sobrecarga no sistema único de saúde fez com que cada vez mais as pessoas procurem serviços privados, com isso o número de beneficiários em operadoras tem apresentado aumento [2]. No meio de saúde suplementar privado, segundo [3] as contratações ocorrem de forma individual pelos indivíduos interessados, denominando-os de planos individuais e familiares ou no formato coletivo, em ambos os cenários os pagamentos ocorrem mensalmente por meio de definições contratuais.

Conforme [4], um fator marcante para o setor da saúde suplementar deve-se pela constituição da Lei nº 9.656/1998 e pela criação do órgão regulador, a Agência Nacional de Saúde (ANS) no ano 2000 [5], em que foi atribuída a responsabilidade de regular o segmento de saúde. Essas medidas modificaram representativamente o setor de saúde, que devido ao aumento da procura por planos privados, originou-se a necessidade de definição de novas regras de proteção aos consumidores, permanência e saída das empresas no mercado, permitindo que ANS realize a promoção do interesse público no que tange a assistência suplementar à saúde, regulando as operadoras e seguradoras relacionadas ao setor e que oferecem planos privados, criando uma relação com os prestadores de serviço e consumidores, mantendo o objetivo de contribuir com o desenvolvimento das ações de saúde [3].

De forma geral, [6] destaca que o setor de saúde suplementar foi regulamentado com o objetivo de unir a garantia assistencial utilizando de padronização e aumento do rol obrigatório de procedimentos, associando a garantia da prestação dos serviços e acompanhando a solvência das operadoras médicas e odontológicas.

No contexto da saúde suplementar um tema que vem ganhando destaque relaciona-se com o setor odontológico, [7] descreve o ramo da odontologia como sendo um serviço de saúde, diferenciando-o de ser apenas uma atividade comercial, sendo mediado pela ética entre o

vínculo do profissional, paciente e o sistema de saúde. Devido sua relevância, em 2012 conforme [8], por meio da Resolução 118, o Conselho Federal de Odontologia (CFO) define a odontologia como uma profissão exercida em prol da saúde, da coletividade, do ser humano e do meio ambiente, sem distinção ou discriminação, contribuindo para que o setor odontológico alcançasse fator de importância assistencial e econômica para o país, principalmente após as regulamentações.

Os modelos assistenciais no que diz respeito ao segmento odontológico, ganharam destaque nas últimas décadas por parte do órgão regulador, considerando principalmente no que se refere as competências atribuídas à ANS observando a Lei 9.656/1998, em seu Art. 8º, reveste-se da maior relevância a proposta do estabelecimento de padrões de qualidade na prestação de serviço de saúde suplementar e continuidade das operadoras odontológicas.

1.1 CONTEXTUALIZAÇÃO DAS OPERADORAS NO BRASIL

Para acompanhamento e gestão econômica das operadoras, o órgão regulador dispõe de algumas ferramentas e processos como por exemplo o Documento de Informações Periódicas das Operadoras de Planos de Assistências de Saúde - DIOPS relatório de envio obrigatório em que as operadoras disponibilizam a ANS informações cadastrais, econômico-financeiras e complementares. Outro relatório solicitado pela agência é o PEONA (Provisão de Eventos Ocorridos e Não Avisados) que disponibiliza de dados contábeis e provisionamento que as operadoras aplicam para gestão financeira.

Relatórios como estes permitem que a agência faça análises econômicas e financeiras das operadoras tanto médicas quanto odontológicas. Segundo dados verificados em [25], em 2017 a ANS publicou em seu portal o Prisma Econômico-Financeiro, referente a consolidação de dados de 2016. O setor fechou o ano de referência com um total de R$161,38 bilhões em contraprestações efetivas faturamento com operação de planos de saúde. O número representa um crescimento nominal em 12,67% quando comparado ao ano de 2015. Por outro lado, as despesas com pagamentos de serviços de assistência à saúde dos beneficiários de planos de saúde tiveram apresentaram valores superiores às contraprestações, resultando no aumento de 14,13% em relação a 2015, totalizando R$137,05 bilhões.

Os resultados apresentados pelo órgão regulador apontaram um fator preocupante para as operadoras, a avaliação sobre os indicadores de sinistralidade, sendo um indicador originado

da relação entre as receitas e despesas assistenciais se é resultada através da taxa de sinistralidade das operadoras. Segundo [10], o número de operadoras odontológicas caiu de 289 para 261 nos últimos três anos, contudo o número de beneficiários apresentou crescimento nos três últimos anos em um ritmo menor, aumentando de 24.201.586 para 26.130.620 descrito na Tabela 1.

Tabela 1 – Número de operadoras odontológicas e beneficiários exclusivamente deste segmento nos anos de 2010 a 2020.

Ano	Quantidade de Operadora	Quantidade de Beneficiários
Dez /2010	374	14.514.074
Dez /2011	369	16.669.935
Dez /2012	359	18.538.837
Dez /2013	344	19.561.930
Dez /2014	343	20.081.836
Dez /2015	327	20.785.367
Dez /2016	305	21.144.527
Dez /2017	291	22.328.276
Dez /2018	289	24.201.586
Dez /2019	280	25.807.205
Set /2020	261	26.130.620

Fonte: ANS 09/2020.

Estes aspectos se tornam fatores preocupantes e que podem influenciar para complicações no setor, desequilíbrio financeiro das operadoras e dificuldade no controle de alguns indicadores básicos de acompanhamento, como por exemplo o sinistro.

1.2 INDICADOR DE SINISTRALIDADE

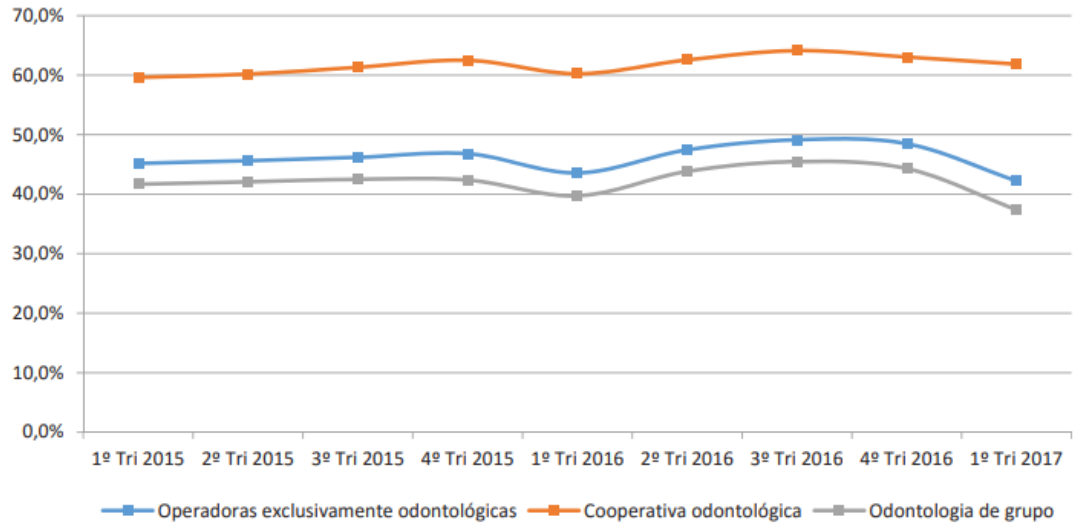
O indicador de sinistralidade, ou sinistro, se dá pela relação entre as despesas e as receitas das operadoras. Segundo [9] essa taxa seria calculada através da divisão dos custos assistenciais versus as receitas assistenciais, ou seja, sinistros pela razão do prêmio. A medida é realizada em porcentagem e se trata de um conceito utilizado nas operadoras como aspecto de gestão e mecanismo para balizar os reajustes contratuais, tendo seu cálculo realizado pela seguinte expressão geral:

$$Sinistralidade = \frac{custos\ assistenciais}{receita\ assistencial} \times 100 \qquad (1)$$

Este indicador possibilita analisar o comportamento da carteira de beneficiários em relação a utilização do plano de saúde ou odontológico, essa análise permite realizar a situação econômico-financeira e se trata de um indicador operacional. Portanto, a sinistralidade é o resultado imediato da utilização dos planos de saúde como por exemplo realização de consultas. Essa utilização de forma descontrolada resulta no aumento dos custos da operadora e consequentemente pode comprometer a gestão financeira da organização.

Segundo dados da Agência Nacional de Saúde [18], nas pesquisas referentes às operadoras odontológicas entre os anos de 2015 a 2017, observa-se na Figura 1 uma estabilidade com relação as receitas em contraposição às despesas advindas da utilização dos beneficiários, o índice de sinistralidade exclusivamente para operadoras odontológicas apresentou uma variação entre 42 e 50, número observado em porcentagem.

Figura 1 – Taxa de sinistralidade das operadoras exclusivamente odontológicas, por modalidade da operadora (2015 a 2017)



Fonte: ANS: Caderno de Informação da Saúde Suplementar - junho/2017.

Diante destas informações, a proposta deste trabalho é avaliar o cenário econômico-financeiro assistencial de uma operadora odontológica de médio porte situada no estado de Minas Gerais, onde foram coletados dados da utilização dos planos entre os anos de 2019 e 2020. Pretende-se ajustar um modelo de regressão linear múltipla representativo ao comportamento dos dados, possibilitando algumas repostas que auxiliem na gestão

administrativa financeira da operadora em estudo. Com base nos resultados da taxa de sinistralidade apresentados pela ANS (Figura 1), a operadora deste estudo estipulou como meta 52% para o indicador operacional de sinistralidade nos anos 2018 a 2020.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 REGRESSÃO LINEAR MÚLTIPLA COM VARIÁVEIS NUMÉRICAS E DUMMYS

Conforme [17], o modelo de regressão linear múltipla é definido pela seguinte expressão matemática:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \epsilon_i \qquad (2)$$

Considere $i, \mathrm{k} = 1, \ldots, n$, sendo n o número que variáveis preditoras selecionadas para o modelo. Seja Y_i a variável resposta observada na i-ésima observação, β_0 se refere ao intercepto do modelo, β_i são os coeficientes que deverão ser associados às variáveis auxiliares ou dependentes x_i e ϵ_i corresponde ao erro aleatório associado à i-ésima observação. Esses componentes contribuem para explicação da variável resposta Y_i. Segundo [17] para a utilização da análise de regressão linear múltipla algumas pressuposições são consideradas sobre o erro a priori:

$$E(\varepsilon_i) = 0 \qquad (3)$$

$$Cov\left(\varepsilon_i, \varepsilon_j\right) = E\left(\varepsilon_i \varepsilon_j\right) - E(\varepsilon_i) E\left(\varepsilon_j\right) = E\left(\varepsilon_i \varepsilon_j\right) = 0, para\ i \neq j, i, j = 1, \ldots, n. \qquad (4)$$

$$Var(\epsilon_i) = \sigma^2 \qquad (5)$$

$$\epsilon_i \sim N(0;\ \sigma^2) \qquad (6)$$

A expressão (3) corresponde a pressuposição de que os erros possuem média nula, a (4) expressa que erro de uma observação é independente do erro de outra observação, a pressuposição (5) indica que a variância do erro é constante e a (6) descreve que os erros são normalmente distribuídos.

Segundo [13] em algumas situações é necessário introduzir, como variável preditora (independente), uma variável categórica no modelo de regressão linear. Nas análises de regressão algumas variáveis qualitativas são codificadas transformando em valores dicotômicos, considerando 0 e 1, denominando-as como variáveis dummy.

2.2 MÉTODO DOS MÍNIMOS QUADRADOS

Segundo [17] o método dos mínimos quadrados objetiva estimar os parâmetros do modelo de regressão linear de modo que a soma dos quadrados dos desvios seja mínima, ou seja, o objetivo é estimar os parâmetros que minimizem o erro, ou seja, estimar os valores de $\beta_0, \beta_1, ..., \beta_k$. Para isso é necessário encontrar as derivas parciais, em relação a cada parâmetro, a função soma de Quadrados dos erros (E).

Inicialmente tem-se o modelo linear múltiplo descrito na seguinte forma:

$$y_{\mathrm{i}} = \beta_0 + \beta_1 x_1 + \cdots + \beta_{\mathrm{k}} x_{\mathrm{i}} \epsilon i \,, \qquad (7)$$

Considere $i, k = 1, \ldots, n$, sendo n as variáveis observadas para o modelo. Tem-se que y_{i} é o valor da variável resposta da i-ésima observação, β_0 e β_1 são parâmetros desconhecidos, x_{i} o valor do preditor da i-ésima observação e ϵi o erro aleatório.

Supondo que existe efetivamente uma relação linear entre as variáveis X e Y, coloca-se a questão de como estimar os parâmetros. Conforme [19], Karl Gauss propôs estimar os parâmetros β_0 e β_{k} visando minimizar a soma dos quadrados dos desvios, ϵi , $i = 1 \ldots n$, sendo este processo denominado como método dos mínimos quadrados. Este método consiste em encontrar os estimadores que minimizam a diferença entre uma observação yi e os valores estimados $\widehat{y_{\mathrm{i}}}$.

$$\epsilon i = y_{\mathrm{i}} - \widehat{y_{\mathrm{i}}} \quad (8)$$

Para isto, seja Q a soma dos n desvios ao quadrado sendo:

$$Q = \sum_{n=1}^{N} (y_{\mathrm{i}} - \beta_0 - \beta_1 x_1 - \cdots - \beta_{\mathrm{k}})^2 \quad (9)$$

De acordo com o método dos mínimos quadrados, as estimativas de β_0 e β_{k} são obtidas quando o critério da função de soma dos quadrados dos erros (E) são mínimos, ou seja, realizando as derivadas parciais em função de β_0 e β_{k}, para encontrar a solução de mínimos quadrados, ou o valor que minimiza a expressão:

$$\frac{\partial \mathrm{E}}{\beta_0} = -2 \sum_{n=1}^{N} (\mathrm{y_i} - \beta_0 - \beta_\mathrm{k} x_\mathrm{i}) \quad (10)$$

$$\frac{\partial \mathrm{E}}{\beta_\mathrm{k}} = -2 \sum_{n=1}^{N} x_\mathrm{i} (\mathrm{y_i} - \beta_0 - \beta_\mathrm{k} x_\mathrm{i}) \quad (11)$$

$$\frac{\partial E}{\partial \beta_0} = 0 \ , \ \frac{\partial E}{\partial \beta_\mathrm{k}} = 0 \qquad (12)$$

Após igualar as expressões (10) e (11) a zero obteve-se a expressão (12). Expandindo os somatórios obter-se:

$$\frac{\partial E}{\partial \beta_0} = \sum_{n=1}^{N} 2 \left(y_\mathrm{n} - (\beta_0 x_\mathrm{n} + \beta_\mathrm{k})\right) . (-x_\mathrm{n}) \ , \qquad (13)$$

$$\frac{\partial E}{\partial \beta_\mathrm{k}} = \sum_{n=1}^{N} 2 \left(y_\mathrm{n} - (x_\mathrm{n} + \beta_\mathrm{k})\right) . (1) \qquad (14)$$

...

$$\frac{\partial E}{\partial \beta_\mathrm{k}} = \sum_{n=1}^{N} 2 \left(y_\mathrm{n} - (x_\mathrm{n} + \beta_\mathrm{k})\right) . (1) \qquad (15)$$

A partir dessas expressões, realiza-se a divisão por dois e aplica-se algumas tratativas algébricas para resultar em uma matriz que contenha os estimadores β que minimizem a soma dos quadrados dos resíduos.

$$\sum_{n=1}^{N} (x_n^2)\, \beta_o + \sum_{n=1}^{N} (x_n) \beta_\mathrm{k} = \sum_{n=1}^{N} x_n y_n \quad , \qquad (16)$$

$$\sum_{n=1}^{N} (x_n)\, \beta_o + \sum_{n=1}^{N} (1) \beta_k = \sum_{n=1}^{N} y_n \qquad (17)$$

Obtendo os valores de β_0e β_k que minimizam o erro e que satisfazem a seguinte equação na forma matricial:

$$\begin{pmatrix} \sum_{n=1}^{N}(x_n^2) & \sum_{n=1}^{N}(x_n) \\ \sum_{n=1}^{N}(x_n) & \sum_{n=1}^{N}(1) \end{pmatrix} \begin{pmatrix} \beta_o \\ \beta_k \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix} \qquad (18)$$

Calculando a inversa da matriz, obtivemos o seguinte resultado:

$$\begin{pmatrix} \beta_o \\ \beta_k \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^{N}(x_n^2) & \sum_{n=1}^{N}(x_n) \\ \sum_{n=1}^{N}(x_n) & \sum_{n=1}^{N}(1) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix} \qquad (19)$$

Assim os valores de melhor ajuste de β_o e β_k são obtidos pela resolução de um sistema linear de equações da matriz anterior. A partir da expressão anterior, ao converter para forma matricial e realizar algumas manipulações das matrizes identificadas, obtêm-se uma matriz que contém os estimadores $\boldsymbol{\beta}$ que minimizem a soma dos quadrados dos resíduos.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad (20)$$

2.3 STEPWISE

Segundo [21] a regressão Stepwise é uma técnica utilizada para eliminar variáveis quando a relação de cada variável preditora com uma variável resposta é testada separadamente para significância estatística, ou seja, auxilia o pesquisador a selecionar as variáveis importantes para o modelo, sendo que as variáveis preditoras que não são significativamente relacionadas à variável resposta são eliminadas ou não consideradas para composição do modelo.

Este método, por sua vez, utiliza o Critério de Informação de *Akaike* (AIC - *Akaike Information Criterion*) na combinação das variáveis dos diversos modelos simulados para selecionar o modelo com melhor ajuste. Quanto menor o AIC, melhor o ajuste do modelo. O AIC é calculado da seguinte forma:

$$AIC = -2\,log(L_P) + 2[(p+1)+1] \qquad (21)$$

em que L_P é definida como a função de verossimilhança e $p + 1$ o número de parâmetros.

Conforme [22] e [23], no método Stepwise podem ser utilizadas as direções:

- *"backward"*: indica "para trás" no que tange a retirada das variáveis, ou seja, quando todos os preditores são incluídos inicialmente e depois são retirados, um a um, até que se identifiquem os melhores preditores.
- *"forward": "pra frente"* indicando inclusão de variáveis a cada passo realizado.
- *"both" ou "stepwise":* descrito como "*ambas"* as direções*,* assemelha-se à regressão stepwise *forward*, mas, ao invés dos preditores serem incluídos individualmente realiza-se a inclusão e exclusão das variáveis preditoras em blocos, ou seja, ocorre uma combinação de seleções para frente e para trás. Inicia-se um modelo sem preditores e adiciona-se sequencialmente os preditores mais contribuintes, após adicionar cada nova variável, as que não foram fornecem melhoria no ajuste do modelo são removidas.

Stepwise é uma modificação da seleção Forward em que cada passo todas as variáveis do modelo são previamente verificadas pelas suas estatísticas F parciais. Uma variável adicionada no modelo no passo anterior pode ser redundante para o modelo devido ao relacionamento com as outras variáveis e, se sua estatística F parcial for menor que um valor pré definido de F com uma determinada significância, remove-se esta variável do modelo.

2.4 TRANSFORMAÇÃO BOX-COX

Segundo [12] nas análises estatísticas em que se busca ajuste de modelo, a pressuposição mais utilizada é a de normalidade, sendo comum normalizar os dados para aproximação da curva normal. No entanto, em muitos casos, os dados apresentam assimetria, ou seja, não possuem aproximação de uma distribuição normal, sendo necessárias transformações nos dados, antes de realizar o ajuste dos modelos.

Conforme observado em [23], um método bastante conhecido e utilizado para transformação dos dados para obter normalidade é a transformação *Box-Cox*, da família paramétrica de transformação potência [2] pode ser escrita pela equação:

$$Y_i^{(\lambda)} = \begin{cases} \frac{{X_i}^{\lambda} - 1}{\lambda}, & para\ \lambda \neq 0 \\ \log(X_i), & para\ \lambda = 0\ e\ y > -c \end{cases} \qquad (22)$$

Com $i = 0,1,2, \dots, n$ e $X_1 \dots, X_n$ os dados iniciais, a transformação de *Box-Cox* consiste em encontrar um valor para λ tal que os dados transformados de Y_i se aproximem de uma distribuição normal. Na expressão (22) pode-se observar que esta transformação tem apenas um parâmetro, o lambda. Se o valor de lambda for igual a zero, realiza-se a transformação logarítmica da sequência inicial, caso o valor de lambda seja difere de zero, a transformação ocorre por lei exponencial. Se o parâmetro lambda for igual a um, a lei de distribuição da sequência inicial permanece inalterada. Segundo [23] dependendo do valor de lambda, a transformação Box-Cox inclui os seguintes casos especiais:

$$\lambda = -1.0, \qquad X_i(\lambda) = \frac{1}{X_i} \qquad (23)$$

$$\lambda = -0.5, \qquad X_i(\lambda) = \frac{1}{\sqrt{X}_i} \qquad (24)$$

$$\lambda = 0.0, \qquad X_i(\lambda) = \ln(X_i) \qquad (25)$$

$$\lambda = 0.5, \qquad X_i(\lambda) = \sqrt{X_i} \qquad (26)$$

$$\lambda = 2.0, \qquad X_i(\lambda) = {X_i}^2 \qquad (27)$$

Segundo [12] após cálculo do lambda que maximiza os resultados e a aplicação desse fator, os valores esperados das observações estarão normalmente distribuídos com variância constante. Conforme [23] para que a lei de distribuição da sequência resultante seja a mais próxima da lei normal quanto possível, o valor ideal do parâmetro lambda deve ser selecionado. Para determinar o valor ideal de lambda deseja-se maximizar o logaritmo da função de verossimilhança, conforme expressão apresentada a seguir:

$$f(x, \lambda) = -\frac{N}{2} \ln \left[\sum_{i=0}^{N-1} \frac{(x_i(\lambda) - \bar{x}(\lambda))^2}{N} \right] + (\lambda - 1) \sum_{i=0}^{N-1} \ln (x_i) \qquad (28)$$

Em que,

$$\bar{x}(\lambda) = \frac{1}{N} \sum_{i=0}^{N-1} x_i(\lambda) \qquad (29)$$

Seja N o número de observações e $i = 1,\dots N-1$. Com base na expressão apresentada, em resumo significa que existe a necessidade de selecionar o parâmetro lambda em que a função de verossimilhança alcance o um valor máximo, ou seja, busca procurar o valor ideal do parâmetro de transformação.

2.5 COEFICIENTE DE CORRELAÇÃO LINEAR

Denominado como coeficiente de correlação de Pearson e sendo atribuída esta nomenclatura devido a Karl Pearson, o responsável pelo desenvolvimento desta estatística. Segundo [24] este coeficiente realiza a mensuração da direção e o gral de relacionamento de duas variáveis quantitativas, ou seja, esta estatística se refere a medida de associação linear entre variáveis de um modelo. Este coeficiente de correlação linear permite quantificar e avaliar a força da relação linear entre as variáveis. Possibilita identificar se esta correlação é positiva, negativa ou se há ausência de relação entre as variáveis.

Segundo [17], sejam X_1, ..., X_i variáveis aleatórias independentes com média amostral representada por $\bar{X}$. A variável dependente é representada por Y_i, com média amostral definida por $\bar{Y}$, $i = 1,2 \dots n$ a i-ésima observação e n o tamanho da amostra. A partir destas definições, o coeficiente linear (r_{xy}) será expresso na seguinte forma:

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} {x_i}^2 {y_i}^2} \qquad (30)$$

Observa-se que $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ e $\bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$, ou seja, r_{xy} para as variáveis X_i e Y_i será o quociente da covariância amostral dessas variáveis pelo produto dos desvios padrão:

$$r_{xy} = \frac{Cov\,(x,y)}{S_x S_y} \qquad (31)$$

O coeficiente de correlação r_{xy} é observado entre os valores de -1 e 1, sendo que o valor zero indica que não há relação linear entre as variáveis, enquanto 1 significa que há uma

correlação positiva indicando que à medida que a variável X_i aumenta a variável resposta Y_i tende a aumentar e -1 indica correlação negativa à medida que a variável X_i, aumenta a variável resposta Y_i tende a diminuir. Conforme [17] quanto mais próximo dos valores extremos, considera-se mais forte a correlação, observando que este coeficiente constitui medida de associação linear.

2.6 MULTICOLINEARIDADE

A capacidade preditiva dos modelos de regressão pode ser prejudicada devido a problemas de multicolinearidade. Segundo [4], "multicolinearidade cria variância "compartilhada" entre variáveis, diminuindo assim a capacidade de prever a medida dependente, bem como averiguar os papéis relativos de cada variável independente".

Neste projeto, para análise estatística da multicolinearidade foi utilizado o teste do fator de inflação da variância (*VIF - Variance Inflation Factor*), função que avalia se as variáveis possuem essa condição. Este teste facilita a verificação do grau que se encontra a relação destas variáveis. O VIF_J é definido pela equação:

$$VIF_J = \frac{1}{1 - {R^2}_{pj}} \qquad (32)$$

em que ${R^2}_{pj}$ será o coeficiente de determinação parcial de X_j em relação às demais covariáveis $(j = 1,2 \ldots, p)$. Para análise aplicação do teste, verifica-se se o fator calculado de cada variável terá valor inferior a 10 representando baixa multicolinearidade [13].

2.7 DISTÂNCIA DE COOKS

A medida denominada distância de Cook [15] é utilizada para análise do ajuste do modelo de regressão, observando o contexto da estimativa de mínimos quadrados. A distância de Cook é calculada removendo o i-ésimo ponto de dados do modelo e recalculando a regressão, ou seja, este método aponta todos os valores no modelo de regressão que mudam quando a i-

ésima observação é removida, visto que o objetivo é avaliar o ajuste do modelo verificando pontos influentes. De modo geral, no contexto de modelos lineares, a matriz:

$$H = X(X'X)X' \quad (33)$$

Conforme [15], H é chamada de matriz chapéu ou de matriz de projeção, visto que esta realiza a transformação dos observados (y_i) sobre os valores ajustados $(\widehat{y_i})$, ou seja, correspondem à matriz de projeção da solução de mínimos quadrados, sendo que os elementos de sua diagonal sejam definidos por h_{ii}, que se referente a medida de "alavancagem". Esta diagonal expressa o quanto as observações estão extremas quando verificadas dentro de um espaço de covariáveis. Pode-se verificar que:

$$\sum_{i=0}^{n} h_{ii} = p \quad (34)$$

Seja alguma observação $i = 1,..,n$ e p um estimador ótimo, sendo n o número de amostras, tal que:

$$h_{ii} > 2.\frac{p}{n} \quad (35)$$

Segundo [15], caso seja constatado que os valores ajustados de h_{ii} apresentam valores superiores conforme expressão (35), então podemos concluir que as observações são pontos de grande "alavancagem", tendo fortes indicativas que o ajuste do modelo não ficou adequado e que existem outliers que interferiram no resultado, ou seja, avalia-se o grau de dependência entre as estimativas $\hat{\beta}$ e cada uma das observações.

Considerando h_{ii} como a intensidade da "alavancagem" exercida pelos valores observados (y_i) sobre os valores ajustados $(\widehat{y_i})$ para todas as observações em um conjunto de dados. Seja o elemento p o número de termos de modelos, pode ser escrito como sendo $\sum_{i=1}^{n} h_{ii}$ e ${r_i}^2$ a medida de discrepância da $i - ésima$ observação, com $i = 1,..,n$. Com base nestas informações, a distância de Cook (D_i) pode ser representada pela seguinte expressão matemática:

$$D_i = \frac{h_{ii}}{p(1 - h_{ii})} r_i^2 \qquad (36)$$

3 MATERIAIS E METODOLOGIA

Os dados utilizados neste trabalho foram coletados junto ao setor que realiza o acompanhamento do sinistro da operadora. Esses dados são referentes aos anos de 2019 e 2020 em que a operadora apresentou piora nos resultados e as análises estavam superficiais sobre os principais fatores que poderia influenciar no sinistro. Além disto o faturamento e despesas foram considerados diretamente para o cálculo do indicador. Diante do exposto foram considerados os dados de junho/2019 a julho/2020 da utilização do plano por beneficiários que estão ligados a empresas do Estado de Minas Gerais.

Foram consideradas 2482 guias referentes a 4 especialidades selecionadas que apresentaram o maior volume de ocorrências. Para análise de regressão outras 8 variáveis preditoras foram agregadas para explicação da variável resposta Y_i que foi medida em porcentagem relacionada aos faturamentos da operadora versus as despesas.

Y_i: Sinistro (SIN)
X_1(R$): valor do procedimento (VALPRO)
X_2(R$): valor real do pago pelo procedimento (VALPA)
X_3 (R$): mensalidade dos beneficiários pagas em 12 meses (MEUSU)
X_4: idade do beneficiário utilizador (IDUTI)
X_5: desvio do valor tabela com o valor pago pelo procedimento (DESP)
X_6: periodicidade (PEROC)
X_7: tipo do utilizador (TIUTI)
X_8: sexo (SEXO)
X_9: Especialidade cirurgia (CIR)
X_{10}: Especialidade consulta/diagnóstico (CON)
X_{11}: Especialidade dentística (DEN)
X_{12}: Especialidade endodontia (ENDO)

As variáveis qualitativas nominais foram transformadas em variáveis dummys conforme indicado na Tabela 2:

Tabela 2 – Descrição das variáveis dummys utilizadas no estudo.

Variável preditora	Atribuição de valores
Periodicidade (PEROC)	Janeiro, fevereiro, julho, junho, dezembro = 1, os demais 0. Sendo os valores com 1 considerados como período de férias
Tipo do utilizador (TIUTI)	Titular = 0 ou Dependente= 1
Sexo	Masculino = 0 ou Feminino = 1
Especialidade cirurgia (CIR)	É cirurgia? Sim = 1 ou Não= 0
Especialidade consulta/diagnóstico (CON)	É consulta? Sim = 1 ou Não= 0
Especialidade dentística (DEN)	É procedimento de dentística? Sim = 1 ou Não= 0
Especialidade endodontia (ENDO)	Assume todos as outras especialidades com o valor 0.

4 RESULTADOS

No início deste trabalho foi abordado o tema de sinistralidade nas operadoras odontológicas e mediante a análises dos dados coletados objetiva-se verificar a influência de variáveis antes não consideradas como fatores relevantes para análise de sinistro.

Inicialmente realizou-se a análise descritiva dos dados separando as variáveis quantitativas contínuas e discretas. De acordo com os dados apresentados na tabela (Tabela 3), se torna possível observar que a variável resposta apresenta um resultado alarmante, sendo valores altíssimo mediante a meta da operadora de 52%, tendo como média calculada o valor de 192,8% valor bastante superior à meta. Além disso, observa-se que o desvio padrão das variáveis VALPA (valor real pago), VAPPRO (valor do procedimento) e MEUSU (mensalidade), apresentam grande desvio em relação à média.

Tabela 3 – Estatísticas descritivas das variáveis quantitativas selecionadas para o modelo.

Variável	Média	Mediana	Máximo	Mínimo	Desvio Padrão
Sinistro (%)	192,80	172,00	277,00	99,00	0,46
X_2 VALPA	43,62	25,50	520,00	5,40	55,24
X_1 VALPRO	123,70	85,00	1300,00	33,00	164,13
X_3 MEUSU	180,10	173,60	278,00	133,20	30,41
X_4 IDUTI	32,89	32,91	70,49	1,82	13,10
X_5DESP	0,17	0,17	1,00	-0,20	0,13

Do total de 2481 guias, 55,26% se tratam da utilização do plano realizada por mulheres e 44,78% refere-se à utilização do plano por homens. Na Tabela 4 são descritos os valores dos custos e receitas separando pela variável preditora, sexo.

Tabela 4 – Relação dos Totais de custo, receita e desvio separadas pela variável sexo.

Totais	Feminino	Masculino
Total Custo	180.842,00	125.961,00
Total Receita	244.284,96	202.611,24
Desvio (Receita – Custo)	63.442,96	76.650,24

Na Tabela 5 são apresentados os valores dos custos por titular, dependente e desvio (em %) em relação ao sexo dos usuários.

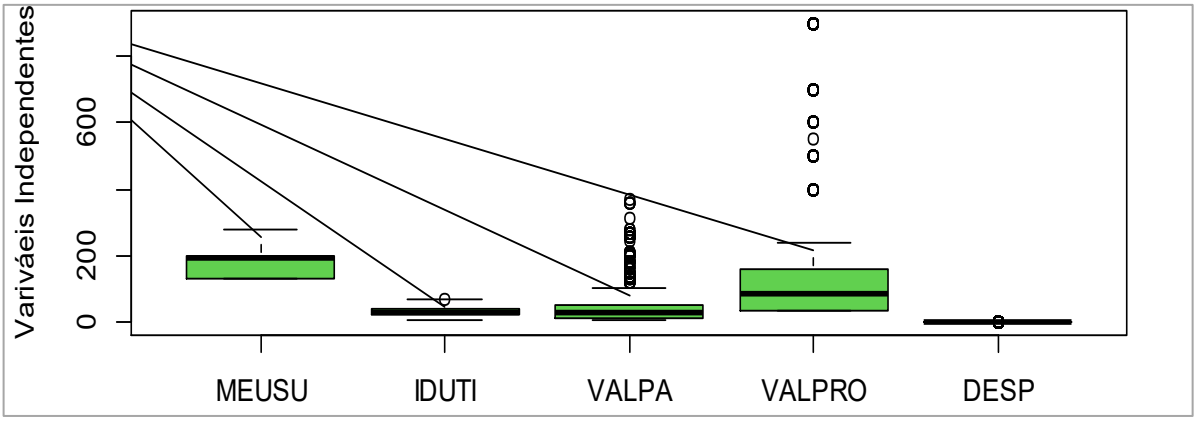
Tabela 5 – Relação dos custos totais por titular, dependente e desvio (em %) em relação ao sexo.

Custos	Feminino	Masculino
Custos Titular	84.762,00	42.736,00
Custos Dependente	95.784,00	83.225,00
Desvio do custo (%)	11,50%	48,65%

É possível notar que tanto para o sexo feminino quando para o sexo masculino, os maiores valores dos custos se relacionam com o tipo de utilizador dependente (Tabela 5), ou seja, os resultados apresentados (Tabela 5) indicam que o sexo feminino possui o maior volume de utilização, seja titular ou dependente. Além disto o desvio do custo indica o quando os titulares utilizam quando comparados aos dependentes, observa-se que o sexo feminino possui apenas 11,50% de diferença entre os custos de dependentes e titulares. Por outro lado, beneficiários do sexo masculino titulares possuem utilização menor do que os dependentes do mesmo sexo.

Utilizando método gráfico (Figura 2) para verificar de maneira informal a simetria das variáveis quantitativas da amostra, verifica-se que as variáveis valor real do pagamento (VALPA) e valor do procedimento (VALPRO) apresentam muitos pontos discrepantes.

Figura 2 – Box-Plot das variáveis quantitativas da amostra.

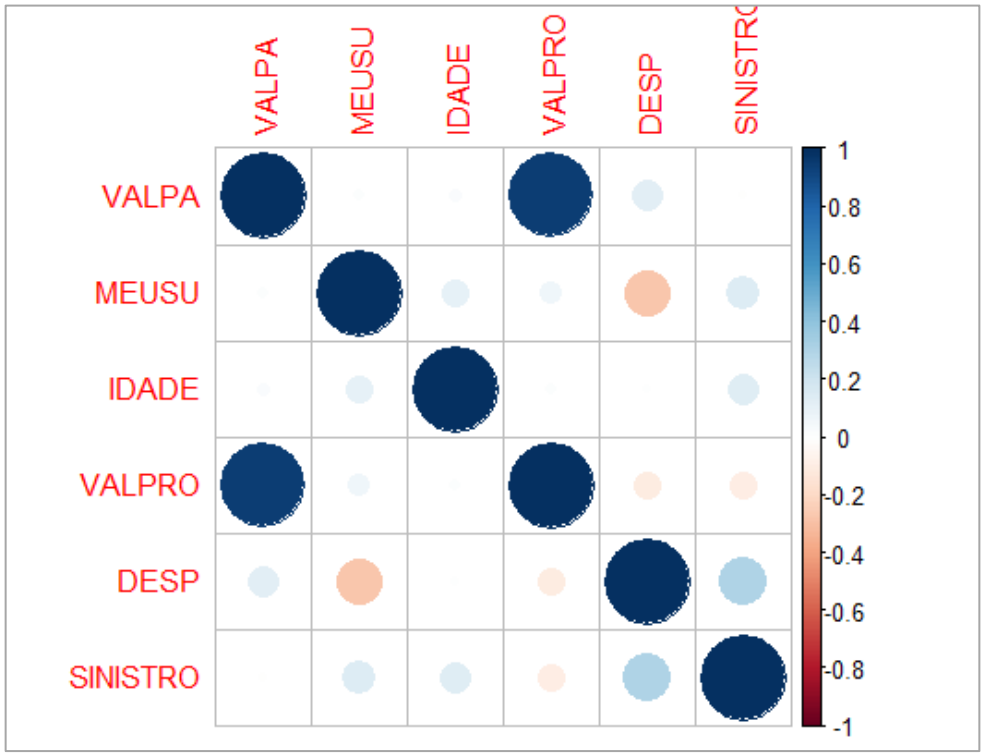


Fonte: Elaborado pelo autor (2020).

Na Figura 3 são apresentadas as correlações entre as variáveis, sendo possível observar associações altas, moderas, baixas e ausência de correlações (nulas). A variável valor do procedimento (VALPRO) com o valor real do pagamento (VALPA) possuem correlação alta e positiva, o desvio do valor tabela com o valor pago (DESP) e mensalidade, possuem correlação moderada negativa, as variáveis desvio do valor tabela com o valor pago (DESP) e valor real

pago (VALPA), mensalidade (MEUSU) e sinistro possuem correlação moderada e positiva. A preditora idade não possui correlação com as variáveis valor real do pagamento (VALPA), valor do procedimento (VALPRO) e desvio da tabela de preços (DESP), situação que se observa para a mensalidade (MEUSU) e a variável VALPA.

Figura 3 – Correlação das variáveis quantitativas do estudo.



Fonte: Elaborado pelo autor (2020).

A variável "sinistro" será a variável dependente e as demais variáveis serão as preditoras selecionadas para análise de regressão. Realizando a análise se variância para verificar se a hipótese (H0) a significância das variáveis preditoras (mensalidade dos beneficiários pagas em 12 meses, periodicidade, tipo do utilizador, sexo, valor real do pago pelo procedimento, valor do procedimento, desvio do valor tabela com o valor pago pelo procedimento, especialidade cirurgia, Especialidade consulta/diagnóstico e Especialidade dentística) sobre o "sinistro". Os valores desse teste podem ser verificados na Tabela 6 e observa-se que as variáveis periodicidade (PEROC), valor real do pago pelo procedimento (VALPA) e Especialidade dentística (DEN) não foram significativas ($p < 0,05$)

Tabela 6 – Análise de variância (ANOVA).

Variáveis	S.Q.	Q.M.	*F value*	*Pr(>F)*
X_3 MEUSU	10,91	10,87	63,18	0,00 ***
X_6 PEROC	0,01	0,02	0,14	0,71
X_7 TIUTI	0,73	0,71	4,15	0,04 *
X_4 IDUTI	7,11	7,05	41,00	0,00 ***
X_8 SEXO	10,41	10,44	60,70	0,00 ***
X_2 VALPA	0,03	0,04	0,22	0,65
X_1 VALPRO	45,45	45,41	263,96	0,00 ***
X_5 DESP	25,79	25,74	149,60	0,00 ***
X_9 CIR	1,47	1,41	8,178	0,04 **
X_{10} CON	14,13	14,07	81,76	0,00***
X_{11} DEN	0,11	0,09	0,50	0,49

S.Q. refere-se a soma dos quadrados da análise se variância e o Q.M refere-se a quadrado médio, *Pr(>F)* refere-se ao valor-p do teste F. A partir desta estatística F e seu valor p abaixo de 0,05, com base em dado estatístico pode-se afirmar ao nível de significância de 5% que variáveis mensalidade (MEUSU), tipo de beneficiário (TIUTI), idade (IDUTI), sexo, valor do procedimento (VALPRO), desvio do valor da tabelado com o valor pago (DESP), especialidade cirurgia (CIR) e especialidade consulta (CON) possuem as médias relacionadas a variável resposta sinistro diferentes entre si.

Para as variáveis período da solicitação (PER), valor real do pagamento (VALPA) e especialidade dentística (DEN) não se rejeita a hipótese nula, visto que apresentaram uma estatística superior a 0,05.

Como apresentado na Tabela 6, foi verificado que quase todas as variáveis foram significativas ao nível de 5% de significância. O período (PEROC), tipo do utilizador (TIUTI), especialidade cirurgia (CIR) e especialidade dentística (DEN) ao nível de significância de 5% não são variáveis significativas para o modelo.

Na tabela 7 pode-se verificar o teste de significância das variáveis preditoras.

Tabela 7 – Teste de significância das variáveis preditoras.

	Coeficiente estimado	Erro Padrão	t value	*Pr(>|t|)*
(*Intercept*)	0,870	0,097	8.976	0,000 ***
X_3 MEUSU	0,003	0,000	11.896	0,000 ***
X_6 PEROC	0,021	0,017	1.197	0,231
X_7 TIUTI	0,030	0,018	1.619	0,106
X_4 IDUTI	0,004	0,001	6.547	0,000 ***
X_8 SEXO	-0,099	0,017	-5.803	0,000 ***
X_2 VALPA	0,010	0,001	2.939	0,003 **

X_1 VALPRO	-0,000	0,000	-2.387	0,017 *
X_5 DESP	1,185	0,094	12.646	0,000 ***
X_9 CIR	-0,024	0,058	-0.423	0,672
X_{10} CON	0,224	0,067	3.328	0,008 ***
X_{11} DEN	0,043	0,061	0.707	0,480

Para o modelo completo verifica-se que o coeficiente de determinação possui valor de 21,38% e o coeficiente de determinação ajustado segue com valor de 21,09%. Em sequência realizou-se o ajuste do modelo utilizando a função de Stepwise com direção "backward, conforme dados apresentados (Tabela 8) em que as estimativas resultaram na expressão (37):

Tabela 8 – Resultados do Modelo Backward Stepwise.

	Estimativa	Erro Padrão	t value	$Pr(>\|t\|)$
(Intercept)	0,8504	0,0725	11,7270	0,0000***
X_3 MEUSU	0,0035	0,0003	11,9530	0,0000***
X_7 TIUTI	0,0284	0,0183	1,5530	0,1206
X_4 IDUTI	0,0045	0,0007	6,6140	0,0000***
X_8 SEXO	-0,0986	0,0170	-5,7800	0,0000***
X_2 VALPA	0,0019	0,0007	2,8960	0,0038
X_1 VALPRO	-0,0005	0,0002	-2,3490	0,0188
X_5 DESP	1,1867	0,0917	12,9350	0,0000***
X_{10} CON	0,2484	0,0392	6,3350	0,0002***
X_{11} DEN	0,0654	0,0340	1,9260	0,0543

$$Y_{Sinistro_{inicial}} = 0{,}8504 - 0{,}0005x_1 + 0{,}0018x_2 + 0{,}0035\,x_3 + 0{,}0045x_4 + 1{,}1867x_5 + 0{,}0284x_7 - 0{,}0986x_8 + 0{,}2484x_{10} + 0{,}0654x_{11} \quad (37)$$

Pode-se observar que as variáveis período de utilização (PEROC), especialidade cirurgia (CIR) e especialidade endodontia (ENDO) não foram significativas para o modelo (Tabela 7).

Posteriormente realizou-se a análise de multicolinearidade utilizando o método de *(VIF)* foram selecionadas apenas as variáveis sexo, idade do utilizador, especialidade consulta, mensalidade paga pelo usuário, desvio do valor tabela de procedimento e especialidade consulta que obtiveram valor menor 10, conforme valores apresentados na Tabela 9.

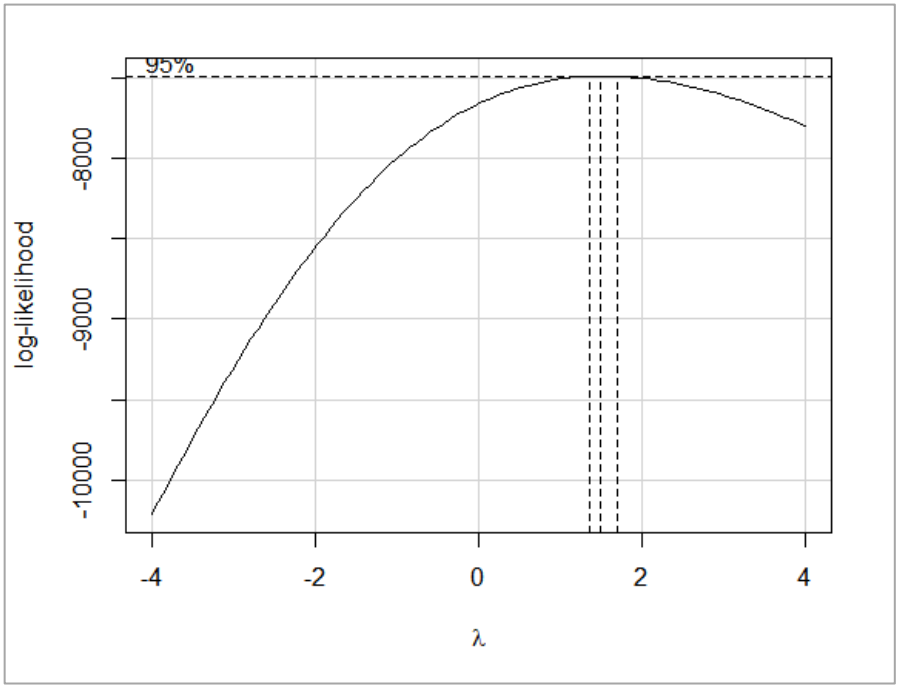
Tabela 9 – Valores VIF das variáveis preditoras.

Variável	*VIF*
Desvio do valor tabelado	2,13
Especialidade Consulta	4,80
Especialidade Dentística	4,12
Idade do utilizador	1,14
Mensalidade	1,66
Sexo	1,04
Tipo Utilizador (titular ou dependente)	1,18
Valor do Procedimento	19,36
Valor Pago	18,65

As variáveis valor do procedimento (VALPRO) e valor real pago (VALPA) apresentaram valores muito superiores a 10, por isso foram desconsideradas para composição do modelo final.

Após a análise de multicolinearidade, verificou-se o pressuposto de normalidade utilizando o teste de Shapiro-Wilk. Ao nível de significância de 10%, devido a não atender ao pressuposto de normalidade, realizou-se a transformação dos dados utilizando o método de transformação box-cox. Na Figura 4 pode-se observar que o valor que maximiza o logaritmo da função de verossimilhança é para $\lambda = 1{,}55$.

Figura 4 – Gráfico de Box e Cox.



Fonte: Elaborado pelo autor (2020).

Após realizar a transformação, obtive-se os seguintes valores descritivos para a variável resposta conforme Tabela 9:

Tabela 10 – Estatística descritiva para o sinistro e sinistro transformado.

Variável	Média	Mediana	Máximo	Mínimo
Sinistro	1,92	1,72	2,77	99,00
Sinistro Transformado	0,62	0,54	1,02	-0,01

Para o modelo final, como apresentado na Tabela 11, foi verificado que quase todas as variáveis foram significativas ao nível de 5% de significância. Somente a variável DEN apresentou um valor p = 0,0762.

Tabela 8 – Teste de significância das variáveis preditores pré selecionadas.

	Coeficiente estimado	Std. Error	t value	*Pr(>|t|)*
(Intercept)	-0,0231	0,0367	-0,6312	0,5295
X_4 IDUTI	0,0022	0,0003	5,9900	0,0000 ***
X_3 MEUSU	0,0023	0,0001	14,2210	0,0000 ***
X_8 SEXO	-0,0525	0,0096	-5,4530	0,0000 ***
X_5 DESP	0,7459	0,0377	19,780	0,0000 ***
X_{10} CON	0,1172	0,0157	7,4600	0,0000 ***
X_{11} DEN	0,0260	0,0146	1,7754	0,0762

Após a transformação da variável do Sinistro, com as variáveis restantes realizou-se novamente o ajuste do modelo regressão com seleção de variáveis. O modelo final na escala transformada foi:

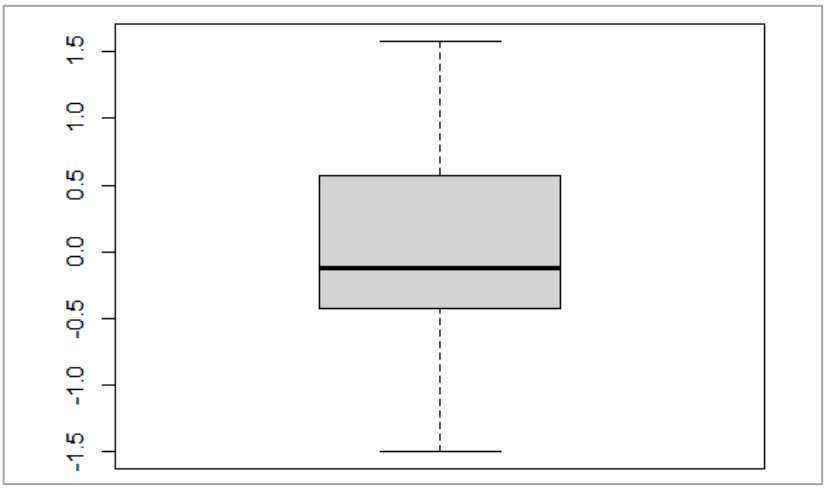
$$Y_{transformada} = -0{,}0231 + 0{,}0023x_3 + 0{,}0022x_4 + 0{,}7459x_5 - 0{,}0525x_8 + 0{,}1172x_{10} + 0{,}0260x_{11} \quad (38)$$

No modelo final ajustado para a escala transformada, o coeficiente 0,0023 associado a variável de mensalidade, indica que para cada unidade que aumentar na mensalidade espera-se que aumente 0,0023 na variável transformada do sinistro gerado, sendo que na escala original corresponde em 0,0193% do sinistro gerado. Para o coeficiente 0,0023, que correspondente a variável idade, indica que a cada ano que aumentar na idade do beneficiário, corresponde ao aumento de 0,0023 no % do sinistro transformado e na escala original corresponde em 0,0187%

do sinistro gerado. E assim para cada variável preditora selecionada, DESP, COM e DEN se relacionam com o aumento da sinistralidade.

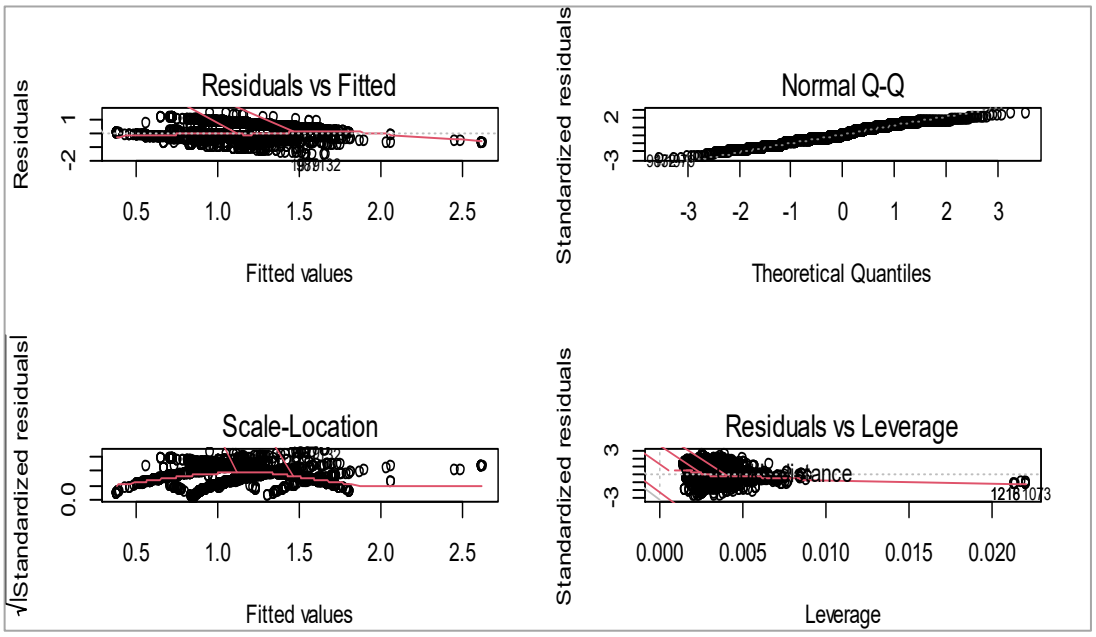
Procedeu-se com a verificação da normalidade dos resíduos. Assim, por ter uma amostra grande e considerando o teorema central do limite, optou-se por testes informais observando que os dados tendem para distribuição normal (Figura 5) e (Figura 6), não foram identificados valores discrepantes.

Figura 5 – Gráfico de Box plot dos resíduos



Fonte: Elaborado pelo autor (2020).

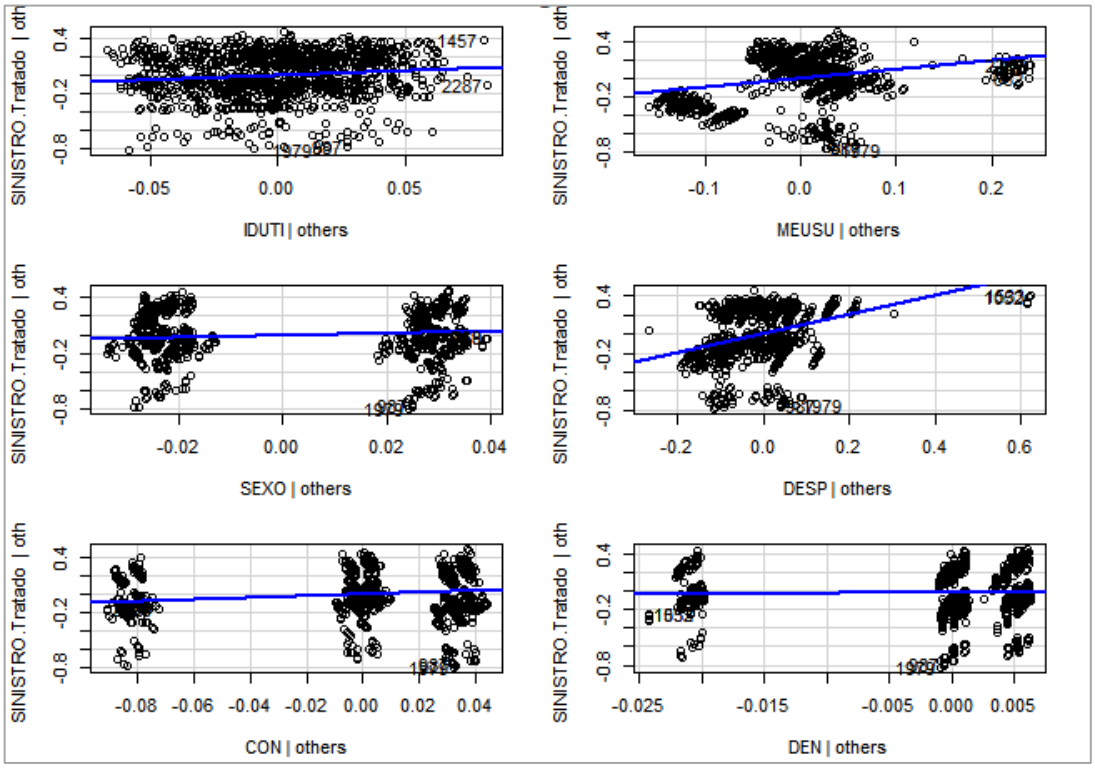
Figura 6 – Gráfico da análise dos resíduos



Fonte: Elaborado pelo autor (2020).

Na Figura 7 apresenta-se outros gráficos para análises de resíduos, denominado gráfico do “leverage Plots” que se refere ao um tipo de gráfico de alavancagem exibindo os resíduos da regressão usando determinada variável desconsiderando as demais, a reta azul indica a regressão o quanto mais distante a observação estiver, maior será o indicativo da falta de ajuste.

Figura 7 – *Leverage Plots*

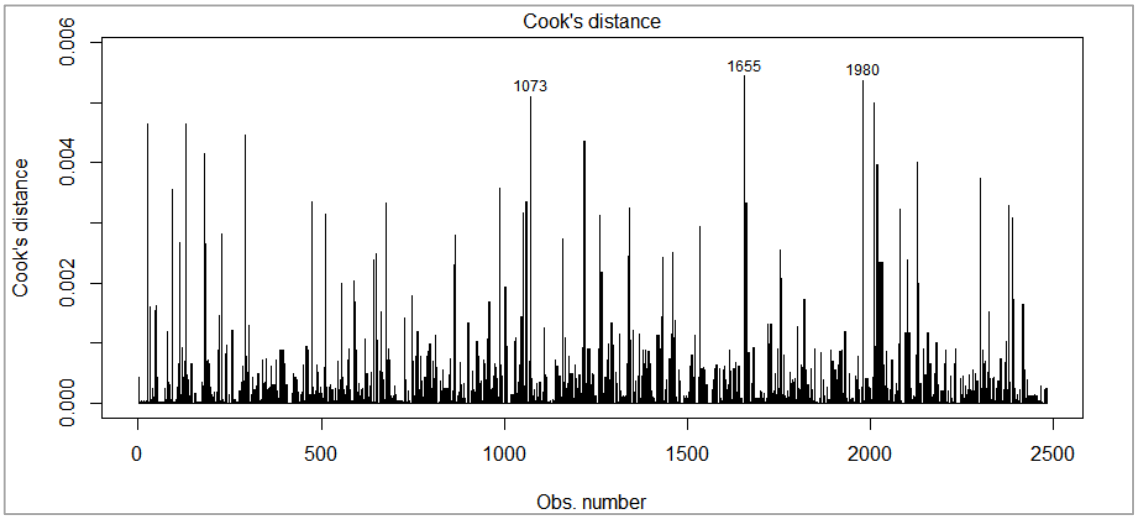


Fonte: Elaborado pelo autor (2020).

Pelos gráficos na Figura 7, tem-se o indicativo que os pontos estão bem próximos a reta e que as distâncias não ultrapassam o valor 1, indicando bom impacto da adição dessas variáveis (IDUTI, MEUSU, SEXO, DESP, COM e DEN) no modelo.

Outro recurso utilizado para verificação do ajuste do modelo foi a distância de Cooks, pela Figura 8. Conforme verificado (Figura 8) não existem indícios de alguma observação verificada no eixo horizontal que esteja acima de 0,5. E se os pontos estivem acima de 0,5 indicaria uma distância grande, ou seja, que seria observações influentes, mas o gráfico apresenta bons resultados e tendo a distância de Cooks muito inferior a 0,5.

Figura 8 – Gráfico da Distância de Cooks



Fonte: Elaborado pelo autor (2020).

O modelo final pode ser escrito na escala original, aplicando a transformação inversa. Assim o modelo é:

$$Y_{Sinistro_{final}} = (0{,}964 + 0{,}0036x_3 + 0{,}0034x_4 + 1{,}16x_5 - 0{,}0814x_8 + 0{,}1817x_{10} + 0{,}0403x_{11})^{0{,}65} \quad (39)$$

Realizando algumas simulações do resultado, para uma pessoa com 33 anos, que paga R$ 14,46 de mensalidade, que fez consulta em um prestador que possui uma tabela com desvio de 0,30 (x_5) comparando com a tabela padrão e que seja mulher, o indicador de sinistro possui um bom resultado, com valor de 76%, acima da meta da operadora. Reduzindo a consulta e o valor do desvio para 0,20 obtém-se o resultado de 51%, abaixo da meta estabelecida pela operadora, apresentando bom resultado.

Após ajuste do modelo os resultados foram considerados pela operadora como relevantes, a variável sexo quando se refere ao público feminino tem maior procura para atendimento consequentemente utiliza mais vezes o plano, então a questão do gênero se torna um fator importante a ser observado.

A preditora idade possui relevância de controle devido à falta de tratamento preventivo, visto que por questões culturais brasileiras fazem com que ocorra o tratamento da doença, não a prevenção e cuidado da saúde, impactando no crescente aumento da utilização conforme idade do beneficiário, quanto mais idade mais vezes o beneficiário procura o plano e os

procedimentos acabam sendo mais caros. A mensalidade possui fator altamente importante porque está ligado diretamente ao faturamento, sendo um dado indispensável para controle dos custos do cupom médio de pagamento.

A variável consulta, sendo uma especialidade voltada para o atendimento inicial ou primário, para os especialistas da operadora em estudo se trata de um fator com necessidade de acompanhamento. Para essa variável indica-se realizar a cobrança de coparticipação seguindo as orientações da ANS para garantia de quatro consultas iniciais sem cobrança, conforme [11].

5 CONCLUSÃO

Com as análises desta monografia, por meio de ferramentas estatísticas foram identificadas seis variáveis que impactam no resultado do sinistro, sendo elas as variáveis mensalidade (x_3), idade do beneficiário (x_4), desvio do valor tabelado com o valor pago pelo procedimento (x_5), sexo (x_8), especialidade consulta/ diagnóstico (x_{10}) e especialidade dentística (x_{11}). Excetuando às variáveis sexo, mensalidade e idade, as demais possuem relação negativa com a variáveis resposta, sendo necessária a redução dos valores associados a estas variáveis para contribuição da redução da sinistralidade.

Faz-se necessário iniciar atividades preventivas para redução do número de consultas e trabalhar com indivíduos observando os grupos por gênero. Outro aspecto importante a ser considerado é a idade dos beneficiários, sugere-se realizar o acompanhamento do sinistro separando em faixas etárias, visto que a operadora em estudo não possui muitos especialistas no estado de Minas Gerais para Odontopediatria fazendo com que quanto menor a idade menor é a utilização neste estado por falta de profissionais. A dentística (x_{11}) por ser uma especialidade mais comum de utilização se torna um fator relevante a considerar para controle de sinistro.

As demais variáveis preditoras não selecionadas, ao nível de significância de 5% não foram significativas para o modelo, não sendo fatores que influenciam estatisticamente na sinistralidade.

6 REFERÊNCIAS

[6] ALVES SL. **Fundamentos, regulação e desafios da saúde suplementar no Brasil**. 1ª ed. Escola, 2015.

[11] **ANS define regras para cobrança de coparticipação e franquia em planos de saúde** - ANS - Agência Nacional de Saúde Suplementar. Disponível em: < http://www.ans.gov.br/aans/noticias-ans/consumidor/4499-ans-define-regras-para-cobranca-de-coparticipacao-e-franquia-em-planos-de-saude >. Acesso em: 14 de dez. de 2020.

[9] ARAÚJO, e SILVA, J.: **Análise de tendência da sinistralidade e impacto na diminuição do número de operadoras de saúde suplementar no Brasil**. Ciência Saúde Coletiva, 23 (10.1590/1413-81232018238.20572016): 2763–2770, 2018. Disponível em: < https://www.researchgate.net/publication/327047794_Analise_de tendencia_da_sinistralidade_e_impacto_na_diminuicao_do_numero_de_operadoras_de_saude_suplementar_no_Brasil >. Acesso em: 30 de nov. de 2020.

[1] BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 1998.

[5] BRASIL. **Lei n.º 9.961/2000**. Dispõe sobre a criação da ANS. Brasília, 2000.

[8] BRASIL. CFO. Conselho Federal de Odontologia. **Código de Ética Odontológica** – Aprovado pela Resolução CFO-118/2012, de 11 de maio de 2012 - Revoga o Código de Ética Odontológica aprovado pela Resolução CFO-42/2003 e aprova outro em substituição. Rio de Janeiro. Disponível em: < http://www.cropr.org.br/uploads/arquivo/724571448d7a83c915ebc18e218042a3.pdf >. Acesso em: 12 de dez. de 2020.

[12] BOX, G. E., & COX, D. R. (1964). ***An analysis of transformations***. *Journal of the Royal Statistical Society*. Series B (*Methodological*), 211-252. Disponível em: < https://www.jstor.org/stable/2984418?seq=1 >. Acesso em: 04 de dez. de 2020.

[22] BRUCE, P. & ANDREW B. ***Practical Statistics for Data Scientists***. O'Reilly Media. 2017

[18] **Caderno de Informação da Saúde Suplementar** - ANS - Agência Nacional de Saúde Suplementar. Disponível em < http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Caderno_informacao_saude_suplementar/caderno_informacao_junho_2017.pdf >. Acesso em: 14 de dez. de 2020.

[15] Cook, R. Dennis. ***Detection of Influential Observations in Linear Regression***. *Technometrics (American Statistical Association),* v. 19, n. 1. p. 15-18,1997.

[4] COSTA, N.R. **O regime regulatório e o mercado de planos de saúde no Brasil**. Ciência e Saúde Coletiva, v. 13, n. 5, p. 1.453-1.462, 2008.

[10] **Dados Gerais** - ANS - Agência Nacional de Saúde Suplementar. Disponível em: < https://www.ans.gov.br/perfil-do-setor/dados-gerais >. Acesso em: 05 de dez. de 2020.

[14] FÁVERO, L. P. L., BELFIORE, P. P., Silva, F. L. d. e Chan, B. L.: **Análise de dados**: modelagem multivariada para tomada de decisões. 2009.

[7] GOMES D, RAMOS FRS. ***The dental professional after productive restructuring****: ethics, the job market and dental public health*. Saúde Soc, 2015; 285-297

[13] HAIR JR., J.F.; WILLIAM, B.; BABIN, B.; ANDERSON, R.E. **Análise multivariada de dados**. 6ª ed. Porto Alegre: Bookman, 2009.

[17] HOFFMANN, Rodolfo - **Análise de regressão: uma introdução à econometria**. São Paulo: Hucitec, 1977.

[23] **JAMES**, G., **WITTEN** D., **HASTIE** T. & **TIBSHIRANI** R., ***An Introduction to Statistical Learning: With Applications in R***. Springer, 1ª ed. 2014.

[21] KEPPEL, G. ***Design and analysis: A researcher's handbook***, 3ª ed. Englewood Cliffs: Prentice-Hall, 1991.

[19] MAROCO, J.; **Análise Estatística – Com utilização do SPSS**, 2ª ed. Lisboa: Silabo, 2003.

[24] MOORE, David S., ***The Basic Practice of Statistics***. *New York: Freeman*. (2007)

[23] Neter, J., Kutner, M.H., Nachtsheim, J., Wasserman, W., ***Applied Linear Statistical Models***,4ª ed. *Richard D. Irwin, Inc*., 1996.

[2] OLIVEIRA DF, KORNIS GEM. **A política de qualificação da saúde suplementar no Brasil**: uma revisão crítica do índice de desempenho da saúde suplementar. *Physis*: Revista de Saúde Coletiva, v.27, n.2, p. 207-231, 2017.

[3] PIETROBON, L.; PRADO, M. L.; CAETANO, J. C. **Saúde suplementar no Brasil**: o papel da Agência Nacional de Saúde Suplementar na regulação do setor. *Physis*, Rio de Janeiro, v. 18, n. 4, p. 767- 783, 2008.

[25] **Prisma econômico – financeiro de saúde suplementar**, ANS. Disponível em: < http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Prisma/2016_prisma_4trim.pdf >. Acesso em: 14 de dez. de 2020.

[16] RAWLINGS, John O.; PANTULA, Sastry G.; DICKEY, David A. **Applied regression analysis: a research tool**. 1998. *Wadsworth and Brooks*.

[20] TABACHNICK, B., & FIDELL, L. S. ***Using multivariate statistics***, 3ª ed. Nova York: Harper Collins, 1996.