

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Especialização em Estatística com Ênfase em Indústria e Mercado**

Daniela Corrêa de Oliveira

**APLICAÇÃO DE REGRESSÃO LINEAR NAS PROJEÇÕES DA CARTEIRA DE**  
**BENEFICIÁRIOS ASSISTIDOS:**  
Estudo de Caso em uma Autogestão de Plano de Saúde

Belo Horizonte  
2019

Daniela Corrêa de Oliveira

**APLICAÇÃO DE REGRESSÃO LINEAR NAS PROJEÇÕES DA CARTEIRA DE  
BENEFICIÁRIOS ASSISTIDOS:**

Estudo de Caso em uma Autogestão de Plano de Saúde

**Versão final**

Monografia de especialização apresentada à Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Estatística, com Ênfase em Indústria e Mercado.

Orientador: Roberto da Costa Quinino

Belo Horizonte  
2019



Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Estatística  
Programa de Pós-Graduação / Especialização  
Av. Pres. Antônio Carlos, 6627 - Pampulha  
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br  
Tel: 3409-5923 – FAX: 3409-5924

## ATA DO 203º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE DANIELA CORRÊA DE OLIVEIRA.

Aos dezessete dias do mês de dezembro de 2019, às 08:00 horas, na Sala 2076 do Instituto de Ciências Exatas, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Daniela Corrêa de Oliveira**, intitulado: **“APLICAÇÃO DE REGRESSÃO LINEAR NAS PROJEÇÕES DA CARTEIRA DE BENEFICIÁRIOS ASSISTIDOS: Estudo de Caso em uma Autogestão de Plano de Saúde”**, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Roberto da Costa Quinino – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 17 de dezembro de 2019.

Prof. Roberto da Costa Quinino (Orientador)  
Departamento de Estatística / UFMG

Prof. Anderson Laécio Galindo Trindade  
DEP / UFMG

Prof. Juliano Ribeiro Moreira  
Instituto de Pesquisas Expertise

## RESUMO

Este trabalho trata da aplicação da técnica estatística de regressão linear como instrumento para projetar a carteira de beneficiários assistidos existentes em uma operadora de autogestão em plano de saúde. Atualmente são diversos os desafios que essas empresas de saúde suplementar sofrem, entre essas dificuldades, cita-se como um ponto de relevância o envelhecimento da população e os altos custos assistenciais. Diante desse cenário, é de extrema importância para a operadora de autogestão obter uma maior assertividade na projeção da carteira de assistidos com a finalidade de definir estratégias para reduzir os índices de sinistralidade da carteira. Para a realização desse estudo utilizou-se dados históricos da carteira de beneficiários assistidos da operadora de autogestão do ramo da saúde suplementar que foram projetados através das técnicas de Regressão Linear.

Palavras-chaves: Regressão linear, Projeção da carteira de assistidos, Operadoras de autogestão em plano de saúde, Saúde suplementar, Envelhecimento da população, Custos assistenciais, Sinistralidade.

## **ABSTRACT**

This work deals with the application of the statistical technique of linear regression as an instrument to project the portfolio of beneficiaries assisted existing in a self-management health plan operator. Currently, there are several challenges that these supplementary health companies face, among these difficulties, the aging of the population and the high healthcare costs are cited as a relevant point. In view of this scenario, it is extremely important for the self-management operator to obtain greater assertiveness in the projection of the assisted portfolio in order to define strategies to reduce the portfolio's loss ratios. To carry out this study, historical data from the portfolio of assisted beneficiaries of the self-management operator in the supplementary health sector were used, which were projected using Linear Regression techniques.

Keywords: Linear Regression, Projection of the assisted portfolio, Health insurance self-management operators, Supplementary health, Aging population, Care costs, Accidents.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> - Dispersão Regressão Linear – Reta de Regressão.	13
<b>Figura 2</b> - Estatística $d$ de Durbin-Watson	30

## LISTA DE TABELAS

<b>Tabela 1</b> – Tabela P-valor	16
<b>Tabela 2</b> – Tabela de decisão - Savin e White	18
<b>Tabela 3</b> – Valores de críticos de d (du e dl) - Savin e White	18
<b>Tabela 4</b> – Descrição das variáveis utilizadas no estudo	20
<b>Tabela 5</b> – Estatísticas descritivas da variável Beneficiários	20
<b>Tabela 6</b> – Teste de Anderson Darling	25
<b>Tabela 7</b> – Parâmetros de estimação do P-valor	26
<b>Tabela 8</b> – Teste de White	27
<b>Tabela 9</b> – Teste de Durbin-Watson	29
<b>Tabela 10</b> – Parâmetros do Teste de Durbin-Watson nível de 95%	29
<b>Tabela 11</b> – Parâmetros do Teste de Durbin-Watson nível de 99%	30
<b>Tabela 12</b> – Teste de Breusch-Godfrey	31

## LISTA DE GRÁFICOS

<b>Gráfico 1</b> – Evolução da Quantidade de Beneficiários Assistidos 2011-2018	21
<b>Gráfico 2</b> – Regressão da Quantidade de Beneficiários Assistidos	23
<b>Gráfico 3</b> – Normalidade dos Resíduos	26



# SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>9</b>
1.1 Objetivo Geral.....	10
1.2 Objetivos Específicos .....	10
<b>2. FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>11</b>
2.1 Envelhecimento da População e Impactos na Saúde Suplementar .....	11
2.2 Regressão Linear Simples .....	12
2.3 Previsão de Novas Observações .....	15
2.4 Teste da Normalidade dos Resíduos – Anderson Darlin.....	15
2.5 Testes de Homocedasticidade/Heterocedasticidade – White.....	16
2.6 Teste Autocorrelação - Durbin-Watson .....	17
<b>3. DESCRIÇÃO DAS VARIÁVEIS E BANCO DE DADOS .....</b>	<b>19</b>
<b>4. ANÁLISE DOS RESULTADOS .....</b>	<b>20</b>
4.1 Estatística descritiva das variáveis.....	20
4.2 Regressão Linear Simples .....	22
4.3 Intervalo de Previsão.....	24
4.4 Teste da Normalidade dos Resíduos .....	25
4.5 Testes de Homocedasticidade/Heterocedasticidade .....	27
4.6 Testes de Autocorrelação .....	28
<b>5. CONSIDERAÇÕES FINAIS.....</b>	<b>32</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>34</b>

## 1. INTRODUÇÃO

Atualmente os desafios na saúde suplementar no Brasil são inúmeros, os custos assistenciais aumentam a cada ano e imensuráveis são os fatores que contribuem para um mercado de risco e abstruso, permitindo muitas dessas operadoras de planos atuarem com dificuldades financeiras. Pode-se mencionar como um dos principais obstáculos da saúde suplementar o envelhecimento da população que acarreta uma maior utilização do plano de saúde na faixa etária que mais gera altos custos, o aumento do rol recorrente dos procedimentos obrigatórios a serem cobertos pelo plano, o aprimoramento das tecnologias que surgem com remédios e equipamentos e cada vez mais onerosos, e outras inúmeras situações enfrentadas pelas operadoras.

Diante um contexto tão conturbado, é de extrema relevância para as operadoras de planos de saúde a elaboração de projeções demográficas futuras que se torna primordial para a orientação de estratégias que necessitem quantificar o público-alvo com a finalidade de contornar a sinistralidade da carteira de beneficiários através do equilíbrio das receitas e custos assistenciais. Esses cenários representam simulações que engloba as dimensões estatísticas, estrutura e distribuição das populações baseadas na análise de tendências históricas e na construção de hipóteses de comportamento futuro para os componentes do crescimento populacional.

A mais atual Projeção de População, divulgada este ano pelo IBGE (Instituto Brasileiro de Geografia e Estatística), estimou demograficamente que a proporção de idosos na população brasileira deverá ultrapassar a de crianças e adolescentes a partir de 2040. Em 2060, conforme o estudo, um quarto da população (25,55%) terá mais de 65 anos. Ou seja, o envelhecimento dos brasileiros é uma tendência irreversível que continuará nas próximas décadas e um ponto de alerta para as operadoras de planos de saúde contornarem os riscos futuros existentes.

Sendo assim, o trabalho proposto visa realizar um estudo de caso em uma operadora de autogestão de plano de saúde de médio porte. Essa operadora possui uma carteira de beneficiários segregados por perfil, ou seja, os participantes titulares

são classificados pelas categorias de empregados ativos, assistidos, ou autopatrocinados.

O foco do estudo será a projeção da carteira de titulares dos beneficiários assistidos, que são os empregados aposentados e pensionistas inscritos na operadora de autogestão. Essa população atualmente representa cerca de 72% (setenta e dois por cento) da carteira de beneficiários titulares total da operadora e sua distribuição de faixa etária se concentra em maior parte na última faixa adotada pela ANS (Agencia Nacional de Saúde Suplementar), no intervalo de 59 (cinquenta e nove) anos ou mais. Conseqüentemente essa carteira também representa o maior índice de sinistralidade, devido à alta utilização dos benefícios ofertados pelo plano de saúde e aos altos custos assistenciais. Dessa forma, o acompanhamento e a previsão semestral dessa carteira de beneficiários é relevante para a Diretoria Executiva aplicar estratégias com intuito de reduzir os índices de sinistralidade, equilibrando as previsões de receitas e custos assistenciais.

Portanto, o intuito do estudo é a aplicação de técnicas estatísticas de regressão linear para projetar essa carteira de beneficiários assistidos existentes nessa operadora de plano de saúde e conseqüentemente subsidiar a operadora, na periodicidade semestral, na tomada de decisões para controle dos riscos e os futuros impactos que ocorrerão na receita e nos custos assistenciais.

### **1.1 Objetivo Geral**

O objetivo da monografia é projetar, através de técnicas estatísticas de Regressão Linear, a carteira de beneficiários assistidos existentes em uma operadora de autogestão em plano de saúde.

De posse dos resultados a operadora poderá redefinir suas estratégias bem como avaliar os impactos e futuros riscos para redução dos índices de sinistralidade.

### **1.2 Objetivos Específicos**

Com o intuito de alcançar o objetivo geral, foram adotados os seguintes procedimentos:

- Projeção da carteira de beneficiários assistidos para: 1º e 2º semestre de 2019;
- Avaliação dos resultados.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### **2.1 Envelhecimento da População e Impactos na Saúde Suplementar**

Os custos com saúde progridem com o envelhecimento da população. De acordo com a Fenasaúde (2018) todo o sistema de Saúde Suplementar se fundamenta no mutualismo (sistema que baseia na diluição de riscos entre um grupo de pessoas) e no denominado pacto intergeracional (sistema pelo qual os mais jovens assumem com parte das despesas dos mais velhos) o que acarreta na maior longevidade da população e possui conflito direto nos planos de saúde.

A contribuição dos beneficiários de planos de assistência médica a partir de 60 anos passou de 23,2%, em dezembro de 2007, para 29,7% em 2017. Destaca-se que nos últimos três anos, a faixa dos beneficiários acima de 80 anos foi a única que apresentou crescimento: 62%. Em todas as demais, houve redução. Em contrapartida, os mais jovens estão deixando o sistema, o que afeta o pacto intergeracional e aumenta em grandes proporções os custos *per capita*.

Segundo com dados do Instituto de Estudos da Saúde Suplementar (IESS) citado pela Fenasaúde (2018), para se ter um conceito do impacto na Saúde Suplementar, um paciente com menos de 18 anos custa em torno ano R\$ 1.500 para seu plano de saúde, enquanto um com mais de 80 pode gerar gastos de R\$ 19 mil por ano. O custo assistencial médio por beneficiário dos 54 aos 58 anos é de R\$ 3.988,23, enquanto que, a partir dos 59 anos, sobe para R\$ 8.036,35 (mais que o dobro).

Outro fato alarmante se trata a mais atual Projeção de População, divulgada este ano pelo IBGE (2018), estimou demograficamente que a proporção de idosos na população brasileira deverá ultrapassar a de crianças e adolescentes a partir de 2040. Em 2060, conforme o estudo, um quarto da população (25,55%) terá mais de 65 anos.

Ou seja, o envelhecimento dos brasileiros é uma tendência irreversível que continuará nas próximas décadas e um ponto de alerta para as operadoras de planos de saúde contornarem os riscos futuros existentes.

A Agência Nacional de Saúde Suplementar (ANS,2018) acrescenta que o custo assistencial de um idoso acima de 60 anos é muito superior ao de um jovem, as pessoas estão vivendo mais, gerando mais custos médico-hospitalares. Conforme a estimativa os gastos com despesas assistenciais até 2030 deverão crescer até 157%, passando de R\$ 149 bilhões para R\$ 383 bilhões.

Para Júnior (2018) com o passar dos anos, os beneficiários de planos de saúde se encontram com reajustes de mensalidade que ultrapassam a inflação.

A percepção é que os custos crescentes, futuramente, acabarão por dificultar o sistema de saúde intermediado pelas operadoras de planos privados, que atinge hoje aproximadamente de 1/4 da população no Brasil.

Os motivos para isso são os mais diversos, como por exemplo: a incorporação de novas tecnologias e medicações de alto custo, passando pelo envelhecimento da população. Ou seja, são várias as influências na evolução do custo assistencial na saúde suplementar. Contudo, quando se observa individualmente esses fatores, verifica-se que o grande vilão da evolução dos preços dos planos é a sinistralidade, que se trata de um índice calculado através da razão entre sinistros realizados (custos de assistência) e o prêmio (receitas da assistência).

A partir desse cenário, é sem dúvida um desafio para as operadoras de planos de saúde contornar essa sinistralidade através de projeções da carteira de beneficiários e equilíbrio dos custos e receitas assistenciais, afim de proporcionar uma melhor assertividade nas tomadas de decisões estratégicas nas precificações dos produtos.

## **2.2 Regressão Linear Simples**

O presente estudo irá aplicar a técnica de Análise de Regressão Linear que é uma das ferramentas estatísticas mais utilizadas na modelagem de dados da carteira de beneficiários assistidos. Trata-se de uma metodologia bastante útil para solução de diversas situações em que precisamos realizar projeções.

A aplicação da técnica de modelagem por regressão linear a uma base de dados resulta da suposição de que a resposta  $Y$  pode ser escrita como combinação linear das variáveis explicativas ( $X$ ). Para Montgomery (2012), quando temos apenas uma variável explicativa a análise de regressão linear é chamada de simples. Esta pode ser representada por:

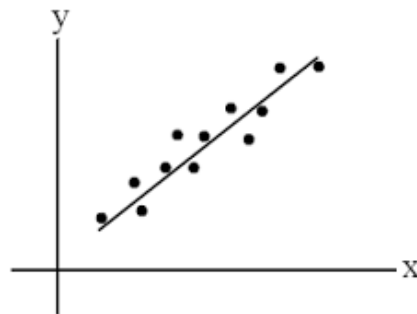
$$E(Y|x) = \beta_0 + \beta_1 x \quad (1)$$

Considerando a interseção  $\beta_0$  e a inclinação  $\beta_1$  coeficientes incógnitos a serem calculados na equação da regressão. Observa-se que cada verificação de  $Y$ , possa ser descrita pelo modelo:

$$Y = \beta_0 + \beta_1 x + \epsilon \quad (2)$$

Destaca-se que  $\epsilon$  é um erro aleatório com média zero e variância  $\sigma^2$ . Adota-se também que os erros são aleatórios e não correlacionados.

Seria bom que as estimativas de  $\beta_0$  e  $\beta_1$  ajustassem uma linha reta em torno das observações como observado na Figura 1.



**Figura 1** - Dispersão Regressão Linear – Reta de Regressão

Sendo assim, o pesquisador Carl Friedrich Gauss (1777-1855), citado por Montgomery (2012), demonstrou que uma boa solução para estimar os parâmetros desconhecidos de uma equação é minimizar a soma dos quadrados dos resíduos.

Esse método, conhecido como mínimos quadrados, consiste em um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma a maximizar o grau de ajuste do modelo aos dados observados. A equação abaixo demonstra como pode-se expressar as  $n$  observações na amostra:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2, 3, \dots, n \quad (3)$$

A diferença entre a observação  $y_i$  e o valor ajustado  $\hat{y}_i$  é denominada de resíduo, ou seja,  $\epsilon_i = y_i - \hat{y}_i$ . O resíduo descreve o erro no ajuste do modelo para  $i$ -ésima observação  $y_i$ .

Logo, a equação para estimativa de mínimos quadrados para interseção e inclinação no modelo de regressão linear simples são:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (4)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \quad (5)$$

em que,

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (6)$$

e

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (7)$$

Em que o coeficiente linear (ou interseção) é representado por  $\hat{\beta}_0$  e o coeficiente angular é representado por  $\hat{\beta}_1$ .

Além dessa estimativa, existe outro parâmetro desconhecido no modelo de regressão que é a variância do termo do erro  $\epsilon$ . A variância do erro,  $\sigma^2$ , reflete a variação aleatória em torno da linha de regressão. Essa variância será aqui estimada através da soma quadrática dos resíduos,  $SQ_E$ :

$$SQ_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (8)$$

$$\hat{\sigma}^2 = \frac{SQ_E}{n-2} \quad (9)$$

Como existem  $n$  pontos experimentais e 2 parâmetros, o número de graus de liberdade, associado ao termo do erro ou resíduo, é igual a  $n - 2$ . Verifica-se que a soma quadrática do erro é a própria expressão da função objetivo do método de mínimos quadrados. Montgomery descreve que uma formula mais simplificada de cálculo  $SQ_E$  pode aplicada pela seguinte equação:

$$SQ_E = SQ_T - \hat{\beta}_1 S_{XY} \quad (10)$$

### 2.3 Previsão de Novas Observações

Uma relevante aplicação da regressão linear é sua capacidade de ser um instrumento para a previsão futuras observações para a variável resposta. Para Montgomery (2012), para um valor especificado do regressor  $x$  é possível prever novas observações para  $Y$ . Caso,  $X_0$  seja o valor desejado do regressor, uma estimativa para o valor da variável resposta  $Y_0$  será:

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta} \quad (11)$$

Um intervalo de previsão para  $Y_0$  é dado por:

$$\hat{y}_0 - t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \leq \hat{Y}_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \quad (12)$$

### 2.4 Teste da Normalidade dos Resíduos – Anderson Darlin

Para utilização do modelo expresso em (2) adota-se que a componente do erro aleatório possua distribuição normal. Na prática é sempre necessário testar se esta hipótese é razoável. O teste Anderson-Darling (STEPHENS, 1974) é frequente para testar se uma amostra de dados deriva de uma determinada distribuição. As medidas de ajuste de aderência submetem-se ao método de estimação aplicado, sendo o teste de Anderson-Darling, é o mais habitual para os métodos de máxima verossimilhança e mínimos quadrados. O teste de Anderson-Darling é um teste aprimorado dos testes de aderência de Chi-quadrado e Kolmogorov- Smirnov. Possui como vantagem um exame mais sensível que os dois mencionados, pois possui maior poder do que os outros dois quando realmente os dados forem normais.

Para constituir um critério de rejeição ou não rejeição do modelo, o teste de Anderson-Darling (1952, 1954) é utilizado para avaliar as seguintes hipóteses:

- $H_0$ : Os dados seguem uma distribuição normal
- $H_1$ : Os dados não seguem uma distribuição normal



A estatística do teste para tomar a decisão é dada por:

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln F(X_i) + \ln(1 - F(X_{n-i+1}))] \quad (13)$$

Em que  $F$  é a função de distribuição acumulada da distribuição específica. Verifica-se que  $Y_i$  são os dados ordenados (NIST, 2002). Os valores críticos ou de rejeição para o teste de Anderson-Darling dependem da distribuição específica que está sendo verificada.

Cabe observar que este teste é ajustado por uma constante, a qual depende do tamanho da amostra ( $n$ ). Estas constantes podem ser localizadas no livro de NIST (2002) ou nos estudos Stephens (1974, 1976 e 1977).

Por fim, após calculado o teste ajustado, conforme tabela abaixo, se  $p$ -valor  $< 5\%$  então não rejeitamos a Normalidade.

**Tabela 1** – Tabela de P-valor

$A^2_m$	P-valor
$A^2_m < 0,200$	$p\text{-valor} = 1 - \exp(-13,436 + 101,14x A^2_m + 223,73x(A^2_m)^2)$
$0,200 < A^2_m < 0,340$	$p\text{-valor} = 1 - \exp(-8,318 + 42,796x A^2_m + 59,938x(A^2_m)^2)$
$0,340 < A^2_m < 0,600$	$p\text{-valor} = \exp(0,9177 - 4,279x A^2_m + 1,38x(A^2_m)^2)$
$A^2_m > 0,600$	$p\text{-valor} = \exp(1,2937 - 5,709x A^2_m + 0,0186x(A^2_m)^2)$

## 2.5 Testes de Homocedasticidade/Heterocedasticidade - White

Segundo Levine et al.(2000) a homocedasticidade demanda que as variações em torno da linha de regressão sejam constantes para todos os valores de  $X$ . Isto expressa que  $Y$  varia na mesma proporção, quando  $X$  for um valor baixo e quando  $X$  for um valor elevado. O contrário desse fenômeno é conhecido por heterocedasticidade, ou seja, a variância não será constante para os valores  $X$  e  $Y$ .

No estudo, será aplicado o teste de White para verificar a existência de heterocedasticidade. De acordo Gujarati e Porter (2011), trata-se de um teste

estatístico otimizado e de simples implementação e consiste primeiramente em realizar a regressão da equação original do modelo:

$$Y = \beta_0 + \beta_1 x_i + \hat{u}_i \quad (14)$$

Em seguida, obtemos os resíduos  $\hat{u}_i$  e realizamos uma segunda regressão auxiliar com a seguinte equação:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + v_i \quad (15)$$

Assim, isso significa que os resíduos ao quadrado da regressão original são calculados por uma nova regressão contra os regressores  $X$  e  $X^2$ . Em seguida, calcula-se o coeficiente de determinação  $r^2$ . Conforme Gujarati e Porter (2011), sob a hipótese nula de que não existe heterocedasticidade, pode-se mostrar que o tamanho da amostra  $n$  multiplicado pelo  $r^2$  da regressão possui distribuição assintótica qui-quadrado com graus de liberdade idênticos ao número de regressores da regressão auxiliar.

Assim:

$$n \times r^2 \sim \chi^2_{gl} \quad (16)$$

Logo, não rejeitaremos a presença de heterocedasticidade se o valor calculado do qui-quadrado exceder o valor crítico no nível de significância escolhido. Senão, será provável concluir que não existe heterocedasticidade, ou que  $\alpha_2 = \alpha_3 = 0$ .

## 2.6 Teste Autocorrelação - Durbin-Watson

Para testar a presença de autocorrelação nos erros de um modelo de regressão em geral usamos o teste de Durbin-Watson. Segundo IMAPE(1998) esse teste existe para identificar a presença de autocorrelação (dependência) nos resíduos de uma análise de regressão. O erro de uma observação qualquer não deve ser influenciado pelo erro de outra observação. Caso exista essa interferência, dizemos que os erros são autocorrelacionados e, apesar dos estimadores ainda serem lineares, não viciados e consistentes, não mais terão a variância mínima.

A autocorrelação pode ser observada por meio da estatística de Durbin-Watson, onde a hipótese básica é a existência de autocorrelação entre resíduos. A estatística de teste pode ser calculada pela fórmula:

$$dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (17)$$

A hipótese nula indica que os erros não são correlacionados e a hipótese alternativa propõe a correlação dos erros. Sendo  $r$  a correlação dos resíduos defasados,  $d = 2(1-r)$ . Caso os erros não são associados, o valor de  $r$  é muito pequeno e  $d$  se aproxima de 2 (dois).

Com o valor verificado da estatística, a tabela de Savin e White (1977) pode ser empregada para uma dada significância, um dado número de observações e de regressores. A tabela fornece valores de críticos de  $d$  ( $d_U$  e  $d_L$ ). O processo de decisão é realizado como:

**Tabela 2** – Tabela de decisão - Savin e White

Hipótese nula	Decisão	Se
Não há autocorrelação positiva	Rejeitar	$0 < d < d_L$
Não há autocorrelação positiva	Sem decisão	$d_L \leq d \leq d_U$
Não há autocorrelação negativa	Rejeitar	$4 - d_L < d < 4$
Não há autocorrelação negativa	Sem decisão	$4 - d_U \leq d \leq 4 - d_L$
Nenhuma autocorrelação, positiva ou negativa	Não rejeitar	$d_U < d < 4 - d_U$

Segue exemplo da tabela de Savin e White (1997):

**Tabela 3** – Valores de críticos de  $d$  ( $d_U$  e  $d_L$ ) - Savin e White

	Nível de significância	Número de variáveis explicativas									
		1		2		3		4		5	
N		$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
	0,01	0,81	1,07	0,7	1,25	0,59	1,46	0,49	1,7	0,39	1,96
15	0,025	0,95	1,23	0,83	1,4	0,71	1,61	0,59	1,84	0,48	2,09

	Nível de significância	Número de variáveis explicativas									
		1		2		3		4		5	
N		$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
	0,05	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
	0,01	0,95	1,15	0,86	1,27	0,77	1,41	0,63	1,57	0,6	1,74
20	0,025	1,08	1,28	0,99	1,41	0,89	1,55	0,79	1,7	0,7	1,87
	0,05	1,2	1,41	1,1	1,54	1	1,68	0,9	1,83	0,79	1,99
	0,01	1,05	1,21	0,98	1,3	0,9	1,41	0,83	1,52	0,75	1,65
25	0,025	1,13	1,34	1,1	1,43	1,02	1,54	0,94	1,65	0,86	1,77
	0,05	1,2	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
	0,01	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
30	0,025	1,25	1,38	1,18	1,46	1,12	1,54	1,05	1,63	0,98	1,73
	0,05	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
	0,01	1,25	1,34	1,2	1,4	1,15	1,46	1,1	1,52	1,05	1,58
40	0,025	1,35	1,45	1,3	1,51	1,25	1,57	1,2	1,63	1,15	1,69
	0,05	1,44	1,54	1,39	1,6	1,34	1,66	1,29	1,72	1,23	1,79
	0,01	1,32	1,4	1,28	1,45	1,24	1,49	1,2	1,54	1,16	1,59
50	0,025	1,42	1,5	1,38	1,54	1,34	1,59	1,3	1,64	1,26	1,69
	0,05	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,7

### 3. DESCRIÇÃO DAS VARIÁVEIS E BANCO DE DADOS

Para a pesquisa tratada nesse trabalho foram utilizados dados da carteira de beneficiários assistidos (aposentados) de uma operadora de plano de saúde, portanto tem-se um estudo de caso. Define-se como estudo de caso a pesquisa “sobre um determinado indivíduo, família, grupo ou comunidade que seja representativo do seu universo, para examinar aspectos variados de sua vida.” (CERVO; BERVIAN, 2004, p. 67).

O período de análise considerado para este estudo será de junho de 2011 a dezembro de 2018, consolidados por semestre. A Tabela 1 apresenta as definições das variáveis estudadas na base de dados.

**Tabela 4** - Descrição das variáveis utilizadas no estudo

<b>VARIÁVEL</b>	<b>CONCEITO</b>
Período	Número sequencial referente a data base que correspondente da informação dos beneficiários no último dia do semestre/ano.
Quantidade de beneficiários Assistidos	A soma total dos beneficiários assistidos (aposentados) existentes.

**Fonte:** Base de dados em estudo

Na base de dados, essas variáveis estão dispostas seguindo a nomenclatura mostrada acima.

#### 4. ANÁLISE DOS RESULTADOS

Nesta seção apresenta-se uma análise descritiva dos dados e os resultados obtidos com a aplicação do método de regressão linear, buscando analisar e demonstrar as principais vantagens oferecidas pela sua aplicação.

##### 4.1 Estatística descritiva das variáveis

Na Tabela 3 encontram-se as algumas estatísticas descritivas da variável estudada.

**Tabela 5** - Estatísticas descritivas da variável Beneficiários

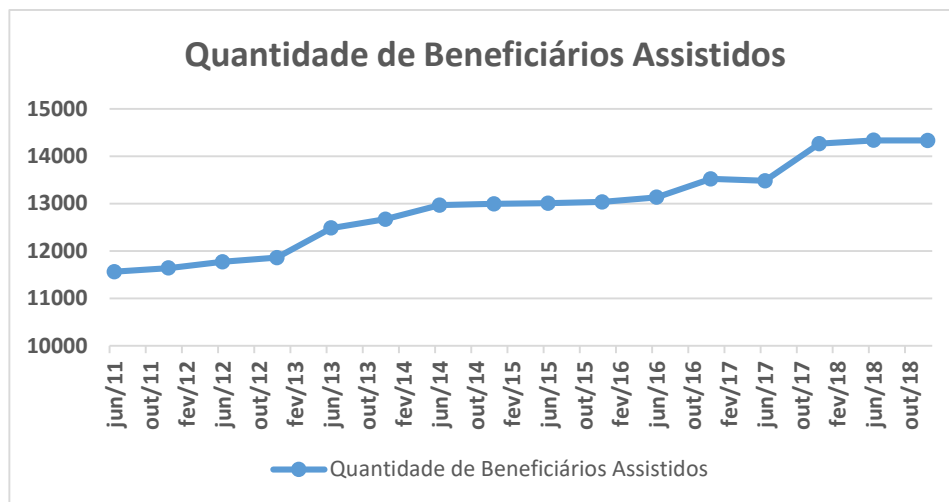
<b>Variável</b>	<b>Quantidade de Beneficiários Assistidos</b>
<b>Média</b>	12.939,69
<b>Erro padrão</b>	230,27
<b>Mediana</b>	13.001,00
<b>Desvio padrão</b>	921,08
<b>Variância da amostra</b>	848.393,70
<b>Curtose</b>	- 0,86
<b>Assimetria</b>	0,03
<b>Intervalo</b>	2.773,00
<b>Mínimo</b>	11.561,00
<b>Máximo</b>	14.334,00
<b>Soma</b>	207.035,00
<b>Contagem</b>	16,00

**Fonte:** Base de dados em estudo

Dessa forma, verifica-se que no período de jun/2011 até dez/2018 o número médio de beneficiários assistidos corresponde a 12.939, sendo que o mínimo foi de 11.561 e o máximo 14.334 beneficiários existentes, o grau de dispersão da amostra apresenta um desvio-padrão de 921,08 e uma mediana de 13.001.

Analisando o perfil de beneficiários assistidos da operadora de plano de saúde observa-se que a carteira de beneficiários está crescendo em uma tendência linear ao longo do tempo. De jun/2011 a dez/2018 houve um crescimento na quantidade de beneficiários de aproximadamente 24%, conforme demonstrado no gráfico 1 abaixo.

**Gráfico 1** – Evolução da Quantidade de Beneficiários Assistidos 2011-2018



O motivo pelo qual ocorre essa tendência no crescimento da população é devido a alteração de categoria do beneficiário. Antes do beneficiário se tornar assistido (aposentado) ele apresentava-se na categoria de empregados em atividade. Logo, com o passar do tempo, o número de empregados em atividade vem reduzindo e o de aposentados aumentando, pois, a empresa patrocinadora não está contratando muitos empregados e terceiriza a mão de obra. Cabe explicar que essa empresa de autogestão em plano de saúde está vinculada a um patrocinador, os beneficiários do plano de saúde são empregados ou aposentados desse patrocinador. Portanto, como o patrocinador não está contratando empregados, os terceirizados não possuem direito ao plano de saúde.

Diante esse ocorrido, verifica-se uma falta de oxigenação na carteira de beneficiários do plano de saúde, pois apenas ocorre uma inversão das categorias de

ativo para assistido. Isso torna-se um desafio para operadora e é de extrema relevância a elaboração de projeções da carteira para a orientação de estratégias que necessitem quantificar o público-alvo com a finalidade de contornar a sinistralidade da carteira de beneficiários através do equilíbrio das receitas e custos assistenciais.

## 4.2 Regressão Linear Simples

Após a análise descritiva dos dados e a identificação de uma tendência linear no crescimento da quantidade de beneficiários assistidos foi realizado uma regressão linear, onde:

X: Período

Y: Quantidade de Beneficiários Assistidos

Logo:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ para } i=1, \dots, n \quad (18)$$

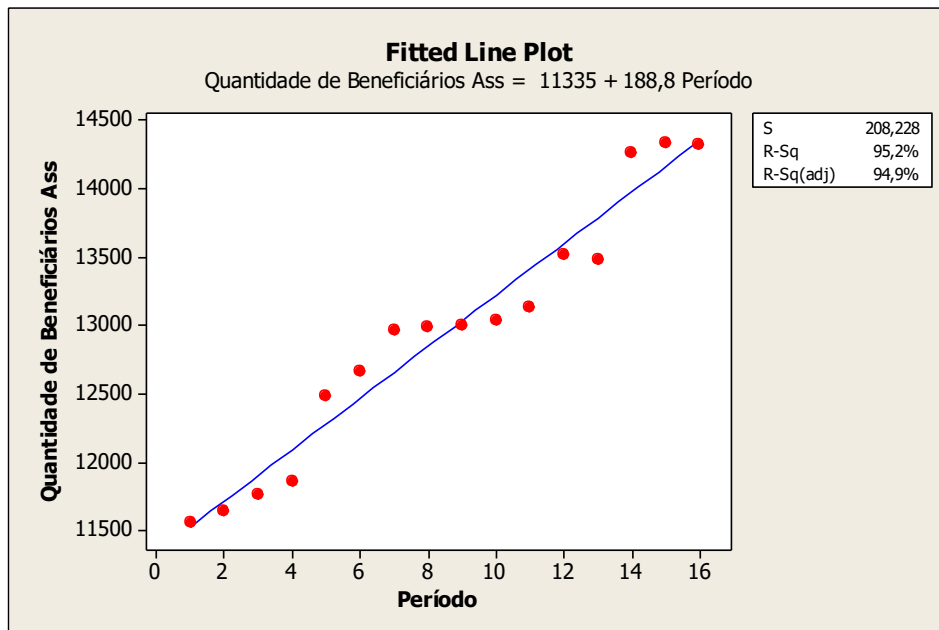
Sendo:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n Y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} = 11.334,9 \quad (19)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - n \bar{x} \bar{Y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = 188,8 \quad (20)$$

Assim, a quantidade de Beneficiários Assistidos estimados = 11.334,9 + 188,8 Período.

**Gráfico 2 –Regressão da Quantidade de Beneficiários Assistidos**



Resultados da regressão:

Preditor	Coef	SE Coef	T	P
Constante	11.334,93	109,2	103,8	0
Período	188,80	11,29	16,72	0

**S = 208,228**    **R-Sq = 95,2%**    **R-Sq(adj) = 94,9%**

Fonte	Soma de quadrados	GL	Quadrado Médio	F <sub>0</sub>
Regressão	$SQR$	p	$QMR = \frac{SQR}{p}$	$F_0 = \frac{QMR}{QME}$
Erro (Resíduo)	$SQE$	n-p-1	$QME = \frac{SQE}{n-p-1}$	
Total	$SQT$	n-1		

Fatores	GL	Soma de quadrados	Quadrado Médio	F	P
Regressão	1	12.118.883	12.118.883	280	-
Erro Residual	14	607.022	43.359		
Total	15	1.272.590			



Diante dos resultados obtidos, avaliou-se a significância da regressão no nível de confiança de 95%, com o seguinte Teste de Hipótese:

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

$$F_{\text{cal}} = 280 \geq F_{\text{tab}} = 6,298$$

Diante a análise, como  $F_{\text{cal}} \geq F_{\text{tab}}$  então rejeita-se  $H_0$  ao nível de significância de 5% e conclui-se que o modelo é razoável. Ou seja, Rejeito  $H_0: B_1=0$ .

### 4.3 Intervalo de Previsão

No estudo elaborado, aplicou-se o intervalo de previsão conforme demonstra a formulação abaixo:

$$\hat{y}_0 - t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \leq \hat{Y}_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]} \quad (21)$$

O intervalo de previsão expressa o erro em prever uma observação futura a partir da distribuição no ponto  $x_0$ .

Logo, aplicou-se o intervalo de confiança de 95% e encontrou-se os seguintes resultados demonstrados abaixo:

<b>Previsões</b>	<b>1º - Semestre/2019:</b>	<b>2º - Semestre/2019:</b>
<b>Previsão (X)</b>	17	18
<b>Y</b>	14.544,4500	14.733,2456
<b><math>t_{\alpha/2; n-2}</math></b>	2,1448	2,1448
<b><math>S_{xx}</math></b>	340	340
<b>Limite Inferior</b>	14.040,1633	14.218,5958
<b>Limite Superior</b>	15.048,7367	15.247,8954

Dessa forma, no nível de significância de 5% é previsto que a quantidade de beneficiários assistidos no 1º-Semestre/2019 corresponda a 14.544 e no 2º-Semestre/2019 infere-se um aumento na quantidade, correspondendo aproximadamente a 14.733.

#### 4.4 Teste da Normalidade dos Resíduos

Para verificar a normalidade dos resíduos, aplicou-se o teste Anderson-Darling. Logo, a estatística do teste para tomar a decisão é dada por:

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln F(X_i) + \ln(1-F(X_{n-i+1}))] \quad (22)$$

Contudo, como a amostra é pequena (16 observações) o valor do AD precisa ser ajustado. O valor AD\* ajustado é dado por:

$$AD' = AD \left( 1 + \frac{.75}{n} + \frac{2.25}{n^2} \right) \quad (23)$$

Considerando o estudo elaborado, os resultados foram:

**Tabela 6 - Teste de Anderson Darling**

Teste de Anderson Darling						
Resíduo Ordenado	F(x <sub>i</sub> )	i	n-i+1	F(x <sub>n-i+1</sub> )	1-F(x <sub>n-i+1</sub> )	AD <sub>i</sub>
-310,268	0,061495	1	16	0,938648	0,061351815	0,348745
-281,676	0,080725	2	15	0,921674	0,078326164	0,949421
-232,107	0,12429	3	14	0,846026	0,153974173	1,236283
-188,881	0,173884	4	13	0,83909	0,160910043	1,564621
-132,312	0,255358	5	12	0,796973	0,203027116	1,664721
-81,472	0,34274	6	11	0,770118	0,229881581	1,74692
-74,516	0,355535	7	10	0,573509	0,426491465	1,532614
-26,085	0,448414	8	9	0,449262	0,55073793	1,311125
-25,654	0,449262	9	8	0,448414	0,551585625	1,482301
37,279	0,573509	10	7	0,355535	0,644464757	1,18194
148,71	0,770118	11	6	0,34274	0,657260152	0,893663
167,141	0,796973	12	5	0,255358	0,744641766	0,750068
199,301	0,83909	13	4	0,173884	0,826115834	0,57259
205,097	0,846026	14	3	0,12429	0,875709546	0,506125
284,937	0,921674	15	2	0,080725	0,919274841	0,300393
310,506	0,938648	16	1	0,061495	0,938504643	0,24564

$$AD = -16+16,2871 = 0,28717 \quad (24)$$

$$AD^* = 0,28717*(1+(0,75/16))+2,25/(16^2) = 0,30313 \quad (25)$$

Sendo  $AD^* = A_m^2$ , seguem os parâmetros de estimação do P-valor:

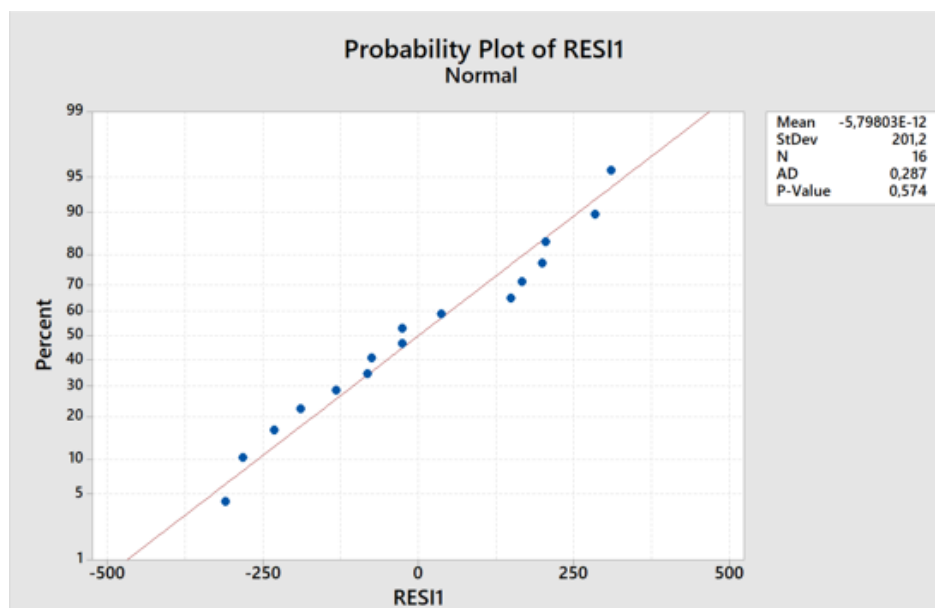
**Tabela 7 – Parâmetros de estimação do P-valor**

$A_m^2$	P-valor
$A_m^2 < 0,200$	$p\text{-valor} = 1 - \exp(-13,436 + 101,14x A_m^2 + 223,73x(A_m^2)^2)$
$0,200 < A_m^2 < 0,340$	$p\text{-valor} = 1 - \exp(-8,318 + 42,796x A_m^2 + 59,938x(A_m^2)^2)$
$0,340 < A_m^2 < 0,600$	$p\text{-valor} = \exp(0,9177 - 4,279x A_m^2 + 1,38x(A_m^2)^2)$
$A_m^2 > 0,600$	$p\text{-valor} = \exp(1,2937 - 5,709x A_m^2 + 0,0186x(A_m^2)^2)$

$$P\text{-Valor} = 1 - \text{EXP}(-8,318 + 42,796 * L30 - 59,938 * (L30^2)) = 0,574 \quad (26)$$

Diante dos resultados, não se rejeita que os dados são normais ao nível de significância 5%. Abaixo segue a demonstração gráfica da normalidade dos resíduos.

**Gráfico 3 – Normalidade dos Resíduos**



#### 4. 5 Testes de Homocedasticidade/Heterocedasticidade

Com o intuito de verificar a homocedasticidade dos dados, aplicou-se o teste de White. Assim, verificou-se os resíduos  $\hat{u}_i$ . Posteriormente, executou-se uma segunda regressão auxiliar com a seguinte equação:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + v_i \quad (27)$$

Isso significa que os resíduos ao quadrado da regressão original são calculados por uma nova regressão contra os regressores  $X$  e  $X^2$ .

Ou seja, considerando a primeira regressão, onde a Inclinação= 188,80 e o Intercepto = 13.334,93, apurou-se os resultados:

**Tabela 8 – Teste de White**

Teste de White						
Período	Quantidade Assistidos	$\hat{Y}$	Resíduos $\hat{u}_i$	Resíduos $\hat{u}_i^2$	Ano <sup>2</sup>	Resíduos <sup>2</sup> estimados
1	11561	11523,72	37,27941	<b>1389,754542</b>	<b>1</b>	3043,451294
2	11638	11712,52	-74,5162	<b>5552,660556</b>	<b>4</b>	13741,29011
3	11769	11901,31	-132,312	<b>17506,40308</b>	<b>9</b>	23143,74741
4	11858	12090,11	-232,107	<b>53873,82329</b>	<b>16</b>	31250,8232
5	12484	12278,9	205,0971	<b>42064,80354</b>	<b>25</b>	38062,51747
6	12667	12467,7	199,3015	<b>39721,07618</b>	<b>36</b>	43578,83023
7	12967	12656,49	310,5059	<b>96413,90298</b>	<b>49</b>	47799,76148
8	12994	12845,29	148,7103	<b>22114,75158</b>	<b>64</b>	50725,31121
9	13008	13034,09	-26,0853	<b>680,4425692</b>	<b>81</b>	52355,47943
10	13034	13222,88	-188,881	<b>35675,98772</b>	<b>100</b>	52690,26614
11	13130	13411,68	-281,676	<b>79341,63408</b>	<b>121</b>	51729,67133
12	13519	13600,47	-81,4721	<b>6637,696369</b>	<b>144</b>	49473,69501
13	13479	13789,27	-310,268	<b>96266,01281</b>	<b>169</b>	45922,33718
14	14263	13978,06	284,9368	<b>81188,95988</b>	<b>196</b>	41075,59783
15	14334	14166,86	167,1412	<b>27936,17287</b>	<b>225</b>	34933,47697
16	14330	14355,65	-25,6544	<b>658,148843</b>	<b>256</b>	27495,9746

$\hat{\beta}_2^*$	$\hat{\beta}_1^*$	$\hat{\beta}_0^*$	R <sup>2</sup>	Estatística de Teste	Valor Tabelado
-647,691	12640,91	-8949,77	18,75%	3,000468661	5,9914

Teste de White:

- $H_0$ : Dados Homocedásticos
- $H_1$ : Dados Heterocedásticos

Logo, através da fórmula:

$$n \times r^2 \sim \chi^2_{gl} \quad (28)$$

Encontrou-se a estatística do teste = 3,000469 e o Valor tabelado = 5,991465. Como a Estatística de Teste é menor que o valor tabelado então não se rejeita  $H_0$  ao nível 95% que os dados são Homocedásticos.

#### 4. 6 Testes de Autocorrelação

Para testar a presença de autocorrelação nos erros de um modelo de regressão, no estudo, aplicou-se o teste de Durbin-Watson. A autocorrelação pode ser observada e designada estatística de Durbin-Watson, onde a hipótese básica é a existência de autocorrelação entre resíduos, que pode ser calculada pela fórmula:

$$dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (29)$$

A hipótese nula indica que os erros não são correlacionados e a hipótese alternativa propõe a correlação dos erros.

Sendo  $r$  a correlação dos resíduos defasados,  $d = 2(1-r)$ . Caso os erros não são associados, o valor de  $r$  é muito pequeno e  $d$  se aproxima de 2 (dois).

Com o valor verificado da estatística, a tabela de Savin e White (1977) pode ser empregada para uma dada significância, um dado número de observações e de regressores. A tabela fornece valores de críticos de  $d$  ( $d_u$  e  $d_l$ ).

Caso  $d$  está entre:

$d < d_l$  ou  $4 - d < d_l$  temos autocorrelação (rejeitando  $H_0$ )

$d > d_u$  ou  $4 - d > d_u$  não temos autocorrelação (falha em rejeitar  $H_0$ )

Ou seja, considerando a primeira regressão, onde a Inclinação= 188,80 e o Intercepto = 13.334,93, apurou-se os resultados:

**Tabela 9 – Teste de Durbin-Watson**

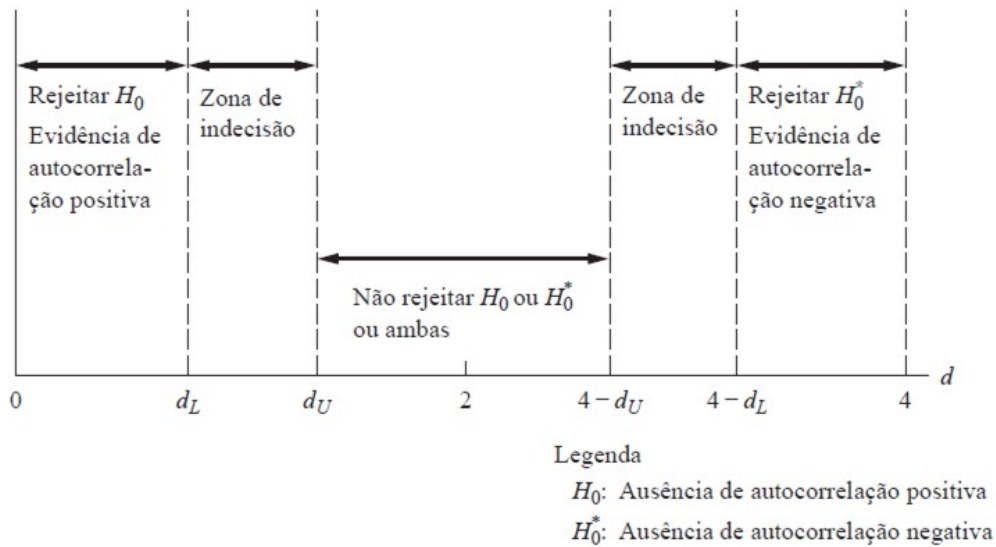
Teste de Durbin-Watson					
Ano	Quantidade Assistidos	$\hat{y}$	Resíduos (ei)	(ei-ei-1)	$e_i^2$
1	11561	11523,72	37,27941	12498,25	1389,755
2	11638	11712,52	-74,5162	3340,33	5552,661
3	11769	11901,31	-132,312	9959,159	17506,4
4	11858	12090,11	-232,107	191147,7	53873,82
5	12484	12278,9	205,0971	33,58884	42064,8
6	12667	12467,7	199,3015	12366,42	39721,08
7	12967	12656,49	310,5059	26177,81	96413,9
8	12994	12845,29	148,7103	30553,5	22114,75
9	13008	13034,09	-26,0853	26502,4	680,4426
10	13034	13222,88	-188,881	8611,021	35675,99
11	13130	13411,68	-281,676	40081,81	79341,63
12	13519	13600,47	-81,4721	52347,42	6637,696
13	13479	13789,27	-310,268	354268,3	96266,01
14	14263	13978,06	284,9368	13875,8	81188,96
15	14334	14166,86	167,1412	37170,14	27936,17
16	14330	14355,65	-25,6544		658,1488
<b>SOMA</b>				<b>818933,6</b>	<b>607022,2</b>

Com os resultados apurados encontrou-se :

$$d = \frac{818933,64}{607022,23} = 1,3490 \quad (30)$$

**Tabela 10 – Parâmetros do Teste de Durbin-Watson nível de 95%**

Parâmetro do Teste (95%)	
0	0
$d_L$	1,106
$d_U$	1,371
$4-d_U$	2,629
$4-d_L$	2,894
4	4



**Figura 2** - Estatística  $d$  de Durbin-Watson

Analisando o resultado obtido, somos levados a uma região de indecisão. No entanto bem próximo do limite. Assim, optou-se em realizar o teste ao nível de 99%.

**Tabela 11** – Parâmetros do Teste de Durbin-Watson nível de 99%

Parâmetro do Teste (99%)	
0	0
dL	0,844
dU	1,086
4-dU	2,914
4-dL	3,156
4	4

Neste caso rejeitamos a presença de Autocorrelação ao nível de 99%. Assim, como no teste ao nível de 95% estávamos próximo ao Limite e no Teste com 99% de confiança não rejeitamos confortavelmente. Então optou-se em não rejeitar a hipótese de não auto correlação aproximadamente com confiança 95%. Adicionalmente realizou-se o Teste de Breusch-Godfrey.

Após definir o modelo e estimar seus coeficientes, o teste de Breusch-Godfrey consiste em efetuar uma regressão do resíduo versus o próprio resíduo defasado no

tempo e as variáveis explicativas. Logo, supondo que o termo do erro siga um processo AR(p) :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t \quad (31)$$

Depois de estimar a equação acima, e definir a ordem de defasagem p, obtém-se a estatística de teste, fundamentada na regressão:

$$\hat{u}_t = \alpha_0 + \alpha_1 x_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \dots + \hat{\rho}_p \hat{u}_{t-p} + v_t \quad (32)$$

Após os resíduos serem estimados, aplicou-se a seguinte equação para verificação da existência de autocorrelação:

$$(n-p) \cdot R^2 = (n-p) \cdot \frac{\sum_{t=1}^{n-p} (\hat{u}_t - \bar{u})^2}{\sum_{t=1}^{n-p} (u_t - \bar{u})^2} \sim X_p^2 \quad (33)$$

Encontrou-se, os seguintes resultados:

**Tabela 12 – Teste de Breusch-Godfrey**

Teste de Breusch-Godfrey			
Resíduos $t$	Resíduos $t-1$	Período $t$	Resíduos $t$ Estimados
-74,516176	37,27941176	2	3,244120088
-132,31176	-74,5161765	3	-32,14585425
-232,10735	-132,311765	4	-50,04182332
205,097059	-232,107353	5	-81,54424093
199,301471	205,0970588	6	60,92150491
310,505882	199,3014706	7	59,87161498
148,710294	310,5058824	8	96,72540312
-26,085294	148,7102941	9	45,13727576
-188,88088	-26,0852941	10	-10,66237138
-281,67647	-188,880882	11	-62,57446179
-81,472059	-281,676471	12	-91,80913798
-310,26765	-81,4720588	13	-26,12263746
284,936765	-310,267647	14	-99,41628987
167,141176	284,9367647	15	94,23561952
-25,654412	167,1411765	16	56,90186682

$\hat{\beta}_2^*$	$\hat{\beta}_1^*$	$\hat{\beta}_0^*$	R <sup>2</sup>	Estatística de Teste	Valor Tabelado
0,827667	0,323963	-10,4884	10,56%	1,583787	3,841459



### Teste de Breusch-Godfrey

- $H_0: \rho = 0$

- $H_1: \rho \neq 0$

Diante dos resultados, como a Estatística do Teste é < Valor Tabelado então não se rejeita  $H_0$  ao nível 95%. Ou seja, não existe evidência de autorrelação ao nível 95%.

## 5. CONSIDERAÇÕES FINAIS

Sabe que a técnica de Análise de Regressão Linear é uma das ferramentas estatísticas mais utilizadas na modelagem de dados, pois trata-se de uma metodologia bastante útil para solução de diversas complicações na estimativa e projeções de dados. Nesse estudo, a aplicação dessa metodologia mostrou-se muito útil para o estudo de caso de uma operadora de plano de saúde projetar sua carteira de beneficiários assistidos.

No estudo, inicialmente, identificou-se que a carteira de beneficiários está crescendo em uma tendência linear ao longo do tempo. De junho/2011 a dezembro/2018 houve um crescimento na quantidade de beneficiários de aproximadamente 24%. Essa informação trata-se de questão crítica e necessita de acompanhamento para operadora, uma vez que se refere a uma carteira envelhecida onde os custos assistências tendem a superar as receitas. Posteriormente, após identificar essa tendência, aplicou-se a regressão linear simples na amostra de dados (beneficiários assistidos de junho/2011 até dezembro/2018). Diante da análise do teste de hipótese sobre regressão, verificou-se que no nível de significância de 5% o modelo proposto era razoável, rejeitando-se  $H_0: B_1=0$ .

Em seguida, foi realizado a previsão da quantidade beneficiário assistidos para o 1º e 2º semestre/2019, encontrando-se as quantidades 14.544 e 14.733, respectivamente. No intuito de testar a normalidade dos resíduos aplicou-se o teste Anderson-Darling, e verificou-se que os dados são normais ao nível de significância de 5%.

Para verificação da homocedasticidade dos dados, aplicou-se o teste de White, como a Estatística de Teste é menor que o valor Tabelado então não se rejeita  $H_0$  ao nível 95% que os dados são Homocedásticos. Por fim, no intuito de analisar a presença de autocorrelação nos erros de regressão, inicialmente aplicou o teste de Durbin-Watson, para os níveis de confiança de 95% e 99%. Assim, como no teste ao nível de 95% estávamos próximo ao Limite e no Teste com 99% de confiança não rejeitamos confortavelmente. Então optou-se em não rejeitar a hipótese de não auto correlação aproximadamente com confiança 95%. Adicionalmente realizou-se o Teste de Breusch-Godfrey. Após definir o modelo e estimar seus coeficientes, esse teste consistiu em efetuar uma regressão do resíduo versus o próprio resíduo defasado no tempo e as variáveis explicativas. Diante dos resultados, como a Estatística do Teste é menor que o valor tabelado então não se rejeitou  $H_0$  ao nível 95%. Ou seja, não existe evidência de autocorrelação ao nível 95%.

Diante todos resultados obtidos e os testes de normalidade, homocedasticidade e autocorrelação, o uso da análise de regressão linear simples foi considerado efetivo.

Para o trabalho em questão, sua utilização conseguiu atender de forma satisfatória a previsibilidade na carteira de assistidos, na periodicidade semestral. Assim, foi fornecido subsídio a operadora, na tomada de decisões para os riscos e os futuros impactos que ocorrerão na receita e nos custos assistenciais devido ao aumento na categoria de beneficiários. Dessa forma, o acompanhamento e a previsão semestral dessa carteira de beneficiários é relevante para a Diretoria Executiva aplicar estratégias com intuito de reduzir os índices de sinistralidade, equilibrando as previsões de receitas e custos assistenciais.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANS – Agencia Nacional de Saúde. Painel de Precificação: **Panos de Saúde 2017**. Disponível em: < file:///C:/Users/Daniela/Downloads/ANS-PAINEL%20DE%20PRECIFICA%C3%87%C3%83O%20PLANOS%20DE%20SA%C3%9ADE%202017.pdf> Acesso em: 10out.2018

CERVO, A.L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2002, 242 p.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Projeção da População 2018: número de habitantes do país deve parar de crescer em 2047**. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/21837-projecao-da-populacao-2018-numero-de-habitantes-do-pais-deve-parar-de-crescer-em-2047>. Acesso em: 20 out.2018.

FenaSaúde. **Envelhecimento da população brasileira impacta em aumento de despesas para os planos de saúde**. Disponível em: <<http://cnseg.org.br/fenasau>de/servicos-apoio/noticias/envelhecimento-da-populacao-brasileira-impacta-em-aumento-de-despesas-para-os-planos-de-saude.html>Acesso em 10 out.2018

IMAPE. **Fundamentos de Avaliações Patrimoniais e Perícias de Engenharia**. 1. ed. São Paulo. Editora Pini. 1998.

JÚNIOR, Antonio Carlos Petto. **Até quando a conta dos planos de saúde ainda vai fechar?** Disponível em: < https://www.conjur.com.br/2018-out-03/antonio-petto-junior-quando-conta-planos-saude-fecha> Acesso em: 10 out. 2018

LEVINE, D.M.; BERENSON, M. L.; STEPHAN, D. Estatística; **Teoria e aplicações**. Rio de Janeiro, LTC, 2000.

MINITAB, Suporte. **Estatísticas de Anderson-Darling em análise de confiabilidade**. Disponível em: < https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/reliability/supporting-topics/distribution-models/anderson-darling-statistics/>. Acesso em 10 nov.2018.

MONTGOMERY, DC; Runger, GC. **Estatística aplicada e probabilidade para engenheiros**. cap.11, p. 263-277. Rio de Janeiro: LTC Editora, 2012 (5ª Edição).

NIST – NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Engineering statistics handbook and handbook of statistical methods**. Sematech: NIST, 2002. Disponível em <http://www.itl.nist.gov/ div898/handbook/,date>. Acesso em: 10 nov.2018.

SAVIN, N. E. e White, K. J. (1977). **The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors**. *Econometrica*, 45(8), p. 1989-1996.

STEPHENS, M.A. EDF: statistics for goodness of fit and some comparisons. **Journal of the American Statistical Association**, v.69, p.730-737, 1974.

STEPHENS, M.A. Asyntotic results for goodness of fit statistics with unkown parameters. **Annals of statistics**, v.4, p.357-369, 1976.

STEPHENS, M.A. Goodness of fit for the extreme value distribution. **Biometrika**, v.64, p.583-588, 1977.