

# PRECIFICAÇÃO ATUARIAL: uma abordagem sensométrica

Pablo Cescon PORTES<sup>1</sup>

Leonardo Henrique COSTA<sup>2</sup>

Eric Batista FERREIRA<sup>3</sup>

<sup>1</sup>Mestre em Estatística Aplicada e Biometria/UNIFAL, Atuário/UNIFAL e-mail: pablo.portes@hotmail.com

<sup>2</sup>Mestre em Ciências Atuariais/PUC-Rio, Universidade Federal de Alfenas - UNIFENAS

<sup>3</sup>Doutor em Estatística e Experimentação Agropecuária/UFLA, Universidade Federal de Alfenas –UNIFAL

**Recebido em: 24/04/2017 - Aprovado em: 11/08/2017 - Disponibilizado em: 30/12/2017**

## RESUMO:

Seguradoras utilizam vários tipos de metodologias para precificação de seguro de carro, modelos de regressão por Poisson, binomial negativa e até mesmo redes neurais. Todas essas modelagens têm como base informações sobre os clientes como: idade, sexo, local de residência, entre outras; além de registros sobre sinistros por parte das seguradoras. Dessa maneira a seguradora admite um risco para cada categoria de pessoa. Porém a percepção de risco que os segurados têm pode ser diferente dos resultados das seguradoras. Sendo assim, objetivou-se com o trabalho, quantificar e analisar a aversão que os clientes de seguro de carro têm aos riscos, com o intuito de auxiliar na avaliação de risco. Para isso, explorou-se o ramo estatístico da Sensometria, que descreve, entende e mede a sensação humana de acordo com estímulos externos. Com os resultados apresentados no trabalho pode-se construir uma nova ferramenta para ser incluída na precificação de seguros de carro.

**Palavras-chave:** percepção dos segurados; seguro de carro; Sensometria.

## ABSTRACT:

Insurance companies use various types of methodologies for car insurance pricing, Poisson or Negative Binomial regression models and even neural networks. All these models take based on customer information such as age, sex, place of residence, among others and records on claims that insurers have, in this way the insurer admits a risk for each category of person. But the pricing of insurance could improve if they used the perception of policyholders about the risks that they run with their cars. Thus, the study aims to quantify and analyze the aversion, that the policyholders, have about risks. To do this, we explored the statistical branch of Sensometric, which describes, understands and measures human feeling according to external stimuli. Therewith, this work will bring a new tool to be included at the car insurance pricing.

**Keywords:** perception of policyholders; car insurance; Sensometric.

## 1. INTRODUÇÃO

Precificações de seguro de carro são feitas de várias formas diferentes, usando vários modelos estatísticos e conjunto de dados. Porém as variáveis, normalmente, são comuns entre as seguradoras de um mesmo país ou região. No Brasil essas variáveis, na maioria das vezes são: se o cliente tem garagem, idade, sexo, modelo do carro, onde mora, onde trabalha, onde estuda, se o cliente

trabalha (ou estuda) em uma cidade e mora em outra, entre outras. Uma boa parte dos modelos de precificação são construídos utilizando métodos de regressão logística, Poisson e binomial negativa, todos eles levando em consideração as variáveis citadas (SILVA; AFONSO, 2015).

Essas modelagens são feitas com base nos bancos de dados das seguradoras, ou seja, informações sobre acidentes e

quantidade de sinistros que já ocorreram para determinada categoria de cliente. Além disso, as seguradoras podem reter informações a *posteriori* dos segurados, por exemplo, número de sinistros que o carro daquele segurado sofreu desde que se iniciou o contrato de seguro.

Seria de grande interesse para as seguradoras analisar o que os segurados pensam sobre o preço que pagam em seus seguros, além da percepção que os mesmos têm sobre os riscos que envolvem seus veículos. O ramo da Estatística que estuda as percepções subjetivas das pessoas sobre fenômenos é a Sensometria.

A Sensometria inicialmente era utilizada na indústria de alimentos, analisando, com base nos cinco sentidos humanos (paladar, tato, visão, olfato e audição) a qualidade de um produto (STORTI et al., 2014). As pessoas que testavam esses produtos poderiam ser provadores profissionais e treinados ou meros consumidores daquele tipo de produto. Porém há trabalhos que aplicam Sensometria em outras áreas, cujo objetivo é quantificar a subjetividade humana de cada indivíduo. Com isso, objetivou-se com o trabalho utilizar a Sensometria para analisar a percepção dos consumidores de seguro sobre os riscos que eles correm.

Utilizou-se um dos métodos estatísticos mais antigos, a Análise de Componentes Principais (ACP). Esse método

é muito utilizado dentro da Sensometria, por se tratar de um método multivariado de análise de dados. Além disso, fez-se uso da aplicação de questionários para averiguar a percepção dos consumidores de seguro de carro. Mas, como foi encontrada uma dificuldade para se estabelecer uma população alvo, os questionários foram aplicados apenas aos professores da Universidade Federal de Alfenas, *campus* Varginha. Dessa forma, a amostra coletada foi satisfatória, tendo em vista que a população era considerada pequena.

Com base nos questionários foi possível transformar as perguntas em variáveis altamente correlacionadas, pois os professores que respondessem alguma das perguntas sendo avessos ao risco tenderiam a responder as outras perguntas de forma coerente, ou seja, também seriam avessos a outros riscos. Dessa forma, seria possível a construção dos componentes principais.

Além disso, utilizou-se de outras variáveis além das perguntas relacionadas ao preço do seguro, como sexo, estado civil, entre outras (com base nos formulários das seguradoras). Essas variáveis foram apenas ilustrativas, pois ajudam a entender melhor o comportamento dos entrevistados.

Foram analisados os gráficos de dispersão dos dados nos eixos dos dois primeiros componentes principais e foram projetados os vetores das variáveis originais e das variáveis ilustrativas quantitativas no

gráfico com os eixos dos dois primeiros componentes principais. Os resultados obtidos no trabalho serão capazes de auxiliar as seguradoras nas precificações de seguros de carro, servindo de base para aumentar ou diminuir o valor do prêmio.

## 2. PRECIFICAÇÃO

Os modelos usados para precificação de seguros de veículos são geralmente, modelos multivariados, que levam em conta uma gama de fatores e variáveis, os quais incidem diretamente no risco de sinistro (acidentes, roubos etc.) e, conseqüentemente, no valor do prêmio pago pelo segurado.

As variáveis usadas pelas seguradoras brasileiras têm como base as informações *a priori*, como: idade, sexo, tipo de carro, tipo de licença, e que podem ser vistas em muitos trabalhos, por exemplo, os artigos de Dionne e Vanasse (1989) e Bailey e Simon (1960), que incluem o local de residência do segurado. Bailey e Simon (1960) colocam também o local de trabalho como variável explicativa do seu modelo de precificação, sendo essa variável também utilizada pelas seguradoras brasileiras. Chang e Fairley (1979) incluem a variável *garagem* em suas modelagens, porém a maioria das seguradoras brasileiras não questiona unicamente se o segurado tem garagem em sua residência, como também se o mesmo estaciona seu veículo na garagem de onde trabalha e de onde estuda.

Após as seguradoras captarem as informações do segurado, elas utilizam-se de métodos estatísticos para determinar o valor do seguro. Normalmente se tem o histórico de acidentes ou roubos para certa idade, sexo ou região e, com isso, usa de métodos como regressão para modelar, por exemplo, o número de acidentes por idade e, assim, apresentar o risco de um segurado de certa idade sofrer acidente.

Dionne e Vanasse (1989) modelam riscos usando regressão pelas distribuições de Poisson e binomial negativa. Jong e Heller (2008) mostram que é possível utilizar modelos de regressão, desde a linear múltipla, até regressões com respostas contínuas, como a regressão por distribuição gamma e gaussiana inversa. O modelo de regressão por Poisson é um dos mais utilizados, pois normalmente os dados vêm em forma de contagem (número de acidentes ou roubos).

Dickison (2005) mostra uma abordagem um pouco diferente para precificação de seguros em geral, usando a teoria da utilidade. A teoria da utilidade é um campo que propicia muitas aplicações, principalmente na área da Economia. A função utilidade, representada por  $u(x)$  descreve uma função que mensura o valor, ou utilidade, que um indivíduo (ou instituição) atribui ao valor monetário  $x$ . Dessa maneira, na ótica das seguradoras, em geral tenta-se maximizar a função utilidade, conforme:

$$u'(x) > 0 \text{ e } u''(x) < 0 \quad (1)$$

Matematicamente, podemos dizer que a seguradora sempre buscará maximizar sua função utilidade, sabendo-se que a mesma tem concavidade voltada para baixo. Todavia essa abordagem é da perspectiva do segurador. Este trabalho também usará o princípio de utilidade, assim, do ponto de vista do segurado, consumidor de seguros, poder-se-á precificar um seguro de carro com base nas preferências do segurado.

### 3. UTILIDADE DOS CONSUMIDORES

De acordo com Pindyck e Rubinfeld (2005), muitas das escolhas que precisam ser feitas pelos consumidores envolvem um considerável grau de incerteza e a maioria vê o risco como algo indesejável. Alguns consumidores consideram-no mais indesejável do que outros. Essa diferença entre as pessoas no modo de agir em relação ao risco está relacionada à utilidade que elas escolhem, a partir das opções entre alternativas de risco.

As pessoas que têm aversão ao risco, que compreende a maior parte da população em geral, estão dispostas a desprender parte da sua renda para evitá-los. Nesse contexto, o valor dos seguros tem como base o prejuízo esperado, ou seja, as pessoas preferem pagar uma quantia certa que seria igual ao prejuízo esperado para a coisa segurada do que pagar o mesmo valor esperado de prejuízo, porém, com uma

variabilidade maior (PINDYCK; RUBINFELD, 2005). Para exemplificar esse cenário será utilizado o seguro de carro, o qual é o foco desse trabalho. As pessoas preferem pagar uma quantia certa de seguro de carro, mesmo que elas saibam que têm uma probabilidade de não bater o carro (prejuízo igual a zero), porque elas sabem que o prejuízo, caso elas venham a sofrer um sinistro, tem a probabilidade de extrapolar o valor do seguro de carro.

Para Rees e Wambach (2008), fazer um seguro é de longe mais racional, no entanto o valor que o consumidor desembolsa para aquisição de um seguro não será usado para produzir nada tangível. Dessa maneira, vemos que o consumidor tem que dar à coisa assegurada um valor utilidade alto para que faça o seguro, além disso, o seguro não pode ser tão alto a ponto de exceder o valor utilidade da coisa. Em geral, os bens de consumo têm valores diferentes para as pessoas, dependendo das circunstâncias nas quais ficaram disponíveis (VARIAN, 2012). O trabalho concentra, justamente, em perceber a utilidade que um grupo de pessoas dá ao seguro de carro, além de averiguar a noção de risco que dessas pessoas, usando técnicas da análise sensorial.

### 4. SENSOMETRIA

Conforme Anzaldúa-Morales (1994), a análise sensorial é um enorme campo da ciência que se ocupa em entender,

descrever, medir e reproduzir os mecanismos de percepção de estímulos externos pelos sentidos humanos básicos. De certa forma, a análise sensorial é o estudo da forma com que o indivíduo sente um objeto ou um estímulo, com base em seus sentidos básicos: olfato, paladar, audição, visão e tato (GARCINDO; FERREIRA, 2013).

No âmbito da indústria de alimentos, por exemplo, essas características são facilmente observadas. Podem ser utilizados o paladar, olfato, visão e tato como variáveis para representar, estatisticamente, um produto, como o café ou um doce. Porém, por se tratarem de variáveis subjetivas, podemos usar as técnicas da Análise Sensorial em outros campos que envolvam pessoas e seus sentidos sobre o mundo externo a elas. Garcindo e Ferreira (2013) utilizaram as técnicas de Sensometria no estudo da percepção de estudantes sobre cursos de Ensino à Distância (EaD). Esse trabalho motivou a aplicação dessa metodologia nas Ciências Atuariais. Um exemplo possível dessa aplicação seria utilizar Análise Sensorial para medir a percepção das pessoas em relação a algum tipo de risco, já na Economia essa metodologia pode ser usada para avaliar como um certo grupo de pessoas percebe um certo fenômeno econômico, como aumento de inflação, por exemplo.

Visto que a forma com que o seguro de carro, ou mais especificamente, o risco de acontecer um sinistro com o carro (roubo,

batida, incêndio, etc.) é percebido pelas pessoas pode ter muitas variáveis. Uma das variáveis, que inclusive é usada pelas seguradoras brasileiras, é a garagem, ou seja, se o segurado tem ou não uma garagem fechada em sua residência. Como será que o consumidor (segurado) percebe essa variável, será que ele dá muito valor a ela, e por isso estaria disposto a desembolsar mais dinheiro em um seguro de carro, caso não tivesse garagem, ou será que o mesmo acha que essa variável não é tão importante quanto outras, como valor do carro, e por isso não estaria disposto a pagar mais pelo seguro mesmo não tendo garagem? É justamente esse modo com que o segurado enxerga os riscos que é o objetivo desse trabalho.

Como a Sensometria está diretamente ligada à análise de dados multivariados, pois a percepção das pessoas em relação aos fenômenos é multidimensional, os métodos utilizados para se quantificar e analisar os dados sensoriais são métodos multivariados, como: análise de arquétipos, análise de agrupamento, análise fatorial, entre outros. Uma das técnicas muito utilizadas é a análise de componentes principais (ACP), que será explicada na próxima seção do trabalho.

## **5. ANÁLISE DE COMPONENTES PRINCIPAIS**

A Análise de Componentes Principais (ACP) é um método multivariado

de modelagem da estrutura de covariâncias, introduzido por Pearson (1901) e, de forma independente, desenvolvido por Hotelling (1933). A ideia dessa técnica é encontrar variáveis latentes, mutuamente não correlacionadas (ortogonais), que são combinações lineares das variáveis estudadas, levando em consideração que essas variáveis são correlacionadas. Assim se o estudo tiver p variáveis tem-se, no máximo, p variáveis latentes. Dessa maneira, ordenam-se essas variáveis latentes de forma que a primeira variável é definida pela combinação linear de maior variância, a segunda terá a combinação linear com segunda maior variância, e assim sucessivamente até a p-ésima variável latente com menor variância. Assim, escolhe-se uma quantidade m de variáveis latentes, que são os componentes principais, de forma que m seja menor que p, para que se tenha uma grande porção total de variância sendo explicada por m componentes principais. É possível, dessa forma, ter uma simplificação da estrutura de covariância do grupo original.

Obtêm-se os componentes principais por meio da diagonalização de matrizes simétricas semipositivas definidas. De acordo com Ferreira (2011), esses cálculos são muito simples pelo fato de existirem inúmeros programas disponíveis que realizam esses cálculos matriciais. Muitas são as aplicações da ACP, como alternativa quando há multicolinearidade na regressão linear múltipla, estimação de fatores, estudos de

divergência, agrupamento entre genótipos em estudo de genética e melhoramento de plantas e animais. Além disso, um dos ramos onde se utiliza muito essa técnica, é no ramo estatístico da Sensometria.

Seja um conjunto de p variáveis  $X_1, X_2, \dots, X_n$ , com médias  $\mu_1, \mu_2, \dots, \mu_p$  e variâncias  $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ , respectivamente. Assumindo que essas variáveis são correlacionadas, existe covariância não-nula entre a i-ésima e a k-ésima variável definida por  $\sigma_{ik}$  com  $i \neq k = 1, 2, \dots, p$ . Assim, pode-se representar estas p variáveis na forma vetorial por  $\mathbf{X} = [X_1, X_2, \dots, X_p]^T$ , com vetor de médias  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^T$  e matriz de variâncias e covariâncias  $\boldsymbol{\Sigma}_{p \times p}$ . Sabendo que os componentes principais  $Y_i$  são combinações lineares das p variáveis, então definimos  $Y_i$  como:

$$Y_i = e_i^T \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p \quad (2)$$

em que o vetor desconhecido  $e_i$  estabelece a i-ésima combinação linear, para  $i=1, 2, \dots, p$ . A variável  $Y_i$ , componente principal, é uma variável latente e, por isso, não mensurada com base na amostra, e sim a partir das p variáveis contidas no vetor  $\mathbf{X}$ . Dessa forma, a ideia é projetar os pontos coordenados originais em um outro plano, maximizando as distâncias entre eles, ou seja, maximizando a variabilidade da variável latente  $Y_i$

(FERREIRA, 2011). Para isso, precisa-se da variância de  $Y_i$  que é dada por:

$$Var(Y_i) = Var(e_i^T \mathbf{X}) = e_i^T Var(\mathbf{X}) e_i = e_i^T \Sigma e_i \quad (3)$$

e a covariância de  $Y_i$  e  $Y_k$  é dada por:

$$Cov(Y_i, Y_k) = e_i^T Var(\mathbf{X}) e_k = e_i^T \Sigma e_k \quad (4)$$

A definição de componentes principais está ligada à maximização da variância de cada um dos componentes. Como não existe o máximo da variância de um componente principal, devemos impor a restrição de que a soma ao quadrado dos coeficientes do vetor, que estabelece a combinação linear dos componentes principais, seja igual à unidade. Dessa forma, maximiza-se a variância do componente principal  $Y_i$ , sujeito à restrição de  $e_i^T e_i = 1$ . Para isso utiliza-se a técnica de multiplicadores de Lagrange, assim:

$$\max[e_i^T \Sigma e_i - \lambda_i (e_i^T e_i - 1)] \quad (5)$$

Com isso, o  $e_i$  é maximizado, sendo  $\lambda_i$  o multiplicador de Lagrange, quando deriva-se essa função em relação a  $e_i$  e iguala-se o sistema de equações a zero, resultando em:

$$(\Sigma - \lambda_i \mathbf{I}) e_i = 0 \quad (6)$$

em que  $\mathbf{I}$  é uma matriz identidade  $p \times p$ . Com essa relação é possível ver que:

$$\Sigma e_i = \lambda_i e_i \quad (7)$$

Por conseguinte, pode-se mostrar que:

$$Var(Y_i) = e_i^T \Sigma e_i = e_i^T \lambda_i e_i = \lambda_i e_i^T e_i = \lambda_i \quad (8)$$

e também que:

$$Cov(Y_i, Y_k) = e_i^T \Sigma e_k = e_i^T \lambda_k e_k = \lambda_k e_i^T e_k = 0 \quad (9)$$

em que,  $i \neq k$ .

De acordo com a equação (9), como já dito anteriormente, os componentes principais são ortogonais entre si, ou seja, a covariância é sempre nula. Além disso, as variâncias dos componentes principais nada mais são do que os autovalores da matriz de variâncias e covariâncias das  $p$  variáveis originais, de forma que as coordenadas dos componentes principais ( $e_i$ ,  $i=1,2,\dots,p$ ) são os autovetores da matriz de variâncias e covariâncias. Como os primeiros componentes principais são os componentes com maior variância, então  $\lambda_1$  é o maior valor, e assim sucessivamente, de forma que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , e assim definindo os componentes principais:  $Y_1 = e_1^T \mathbf{X}, \dots, Y_p = e_p^T \mathbf{X}$ , respectivamente.

Utilizando a decomposição espectral proposta por Johnson e Wichern (2007), que é dada por:

$$\Sigma = \mathbf{P} \Lambda \mathbf{P}^T \quad (10)$$

em que  $\mathbf{P}$  é a matriz composta pelos autovetores de  $\Sigma$  em suas colunas, e  $\Lambda$  é a matriz diagonal de autovalores de  $\Sigma$ , ou seja, pode-se observar que:

$$\begin{aligned} tr(\Sigma) &= tr(\mathbf{P} \Lambda \mathbf{P}^T) = tr(\Lambda \mathbf{P} \mathbf{P}^T) = \\ &= tr(\Lambda \mathbf{I}) = tr(\Lambda) = \sum_{i=1}^p \lambda_i \quad (11) \end{aligned}$$

Porém,  $tr(\Sigma)$  também pode ser escrito como:

$$tr(\Sigma) = \sum_{i=1}^p \sigma_{ii} \quad (12)$$

Dessa forma, conclui-se que:

$$\sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i \quad (13)$$

Isso significa que a variância total contida nas variáveis originais é igual a variância total contida nos componentes principais. Quando se quer obter  $k$  componentes principais de forma que  $k < p$ , utiliza-se esse critério para entender quanto da variabilidade total das variáveis originais estão sendo explicadas por cada componente principal. Assim, procura-se o menor número de componentes principais que expliquem a maior variabilidade dos dados. Graficamente, a visualização fica mais fácil de ser entendida quando há dois componentes principais, os quais seriam os eixos (abscissa e ordenada) do gráfico.

## 5.1 VARIÁVEIS SUPLEMENTARES QUANTITATIVAS

As variáveis suplementares (ou ilustrativas) são diferentes das variáveis ativas. De acordo com Hussonet *et. al.* (2011), por definição, variáveis ativas contribuem para a construção dos componentes principais, ao contrário das variáveis suplementares, que apenas ajudam a entender as variáveis ativas de forma adicional aos componentes principais, por isso também são chamadas de variáveis ilustrativas.

Existem dois tipos de variáveis suplementares, as quantitativas e as qualitativas. As variáveis quantitativas são representadas da mesma forma que as variáveis ativas, para auxiliar na interpretação gráfica dos componentes principais. As coordenadas das  $k'$  variáveis suplementares quantitativas no componente  $Y_i$  correspondem à correlação entre as  $k'$  variáveis e o  $Y_i$  componente principal. Mais formalmente, para se calcular as coordenadas das  $k'$  variáveis suplementares no componente  $Y_i$ , utiliza-se a seguinte fórmula:

$$G_{Y_i}(k') = \frac{1}{\sqrt{\lambda_{Y_i}}} \quad (14)$$

em que  $G_{Y_i}(k')$  são as coordenadas da variável  $k'$  no eixo do componente principal  $Y_i$  e  $\lambda_{Y_i}$  é o autovalor associado ao componente  $Y_i$ . Dessa forma, as variáveis suplementares não entram nos cálculos dos componentes principais, mas entram graficamente junto com as variáveis ativas.

## 5.2 VARIÁVEIS SUPLEMENTARES QUALITATIVAS

As variáveis suplementares qualitativas não podem ser representadas da mesma forma que as variáveis quantitativas, por se tratarem de categorias. Por isso, elas são puramente ilustrativas, pois não tem como calcular as coordenadas dessas variáveis não numéricas. Dessa forma, as variáveis

suplementares qualitativas (ou categóricas) são representadas como o baricentro das variáveis ativas que pertencem àquela variável suplementar.

O conceito de baricentro, de acordo com o dicionário *online* Michaelis (2015) é centro de gravidade, ou seja, é um ponto em torno do qual existe um equilíbrio de forças. Ou seja, o baricentro de uma nuvem de dados é o ponto que tenta sofrer influência geométrica de todos os pontos, tentando ficar “no meio” da nuvem.

Graficamente, quando se projeta a nuvem de dados nos dois primeiros componentes principais, tem-se um gráfico com dois eixos, assim, o baricentro de cada uma das categorias dos dados são as variáveis suplementares qualitativas. Com isso, pode-se visualizar como os dados estão se comportando de acordo com aquelas variáveis suplementares qualitativas.

Dessa maneira tem-se que o objetivo do trabalho foi quantificar as percepções humanas sobre os riscos que envolvem automóveis, com base na quantidade de dinheiro que as pessoas estão dispostas a desembolsar nos seguros de carro. Para isso, utilizou-se a Sensometria, ramo da Estatística que analisa e descreve as sensações humanas. Os resultados que foram obtidos no trabalho devem ser capazes de auxiliar seguradoras e atuários na precificação de seguros. Para isso utilizou-se ACP, uma das técnicas usadas na Sensometria, para estudar o comportamento

das pessoas em relação ao seguro de carro, além disso, compararam-se os riscos analisados no trabalho com os cálculos feitos pelas seguradoras para os mesmos riscos.

## **6. MATERIAL E MÉTODOS**

### **6.1 DADOS**

Para analisar as percepções que os segurados têm com relação aos riscos que correm envolvendo seus carros, uma das maneiras de captar isso é através de questionários. Podem ser feitas perguntas que instiguem o senso de risco que a pessoa tem com base em suas experiências passadas, ou seja, simular situações que coloquem o entrevistado em uma situação desconfortável financeiramente. Um exemplo disso é perguntar para o segurado que possui garagem em sua residência, quanto ele estaria disposto a pagar a mais caso não tivesse garagem, com isso captamos a ideia que a pessoa tem sobre o quanto a garagem interfere no risco de seu carro ser roubado.

Para poder quantificar essas perguntas subjetivas foi feita uma escala de 0% a 100% do preço atual pago pelo segurado, ou seja, em cada pergunta o entrevistado respondeu quanto ele pagaria a mais no seguro do seu carro em porcentagem. Dessa forma, cria-se uma variável para cada pergunta respondida pelos entrevistados. Por exemplo, se o entrevistado estivesse disposto

a pagar 50% a mais, com base no valor do seu seguro atual, por, hipoteticamente, não possuir garagem, ele marcaria a opção “de 40% a 50%”.

Os questionários foram aplicados para 36 professores da Universidade Federal de Alfenas, *campus* Varginha, entre os meses de fevereiro de 2015 e julho de 2015.

## 6.2 MÉTODOS

Para se obter a estimativa da proporção de professores que responderam alguma das perguntas foi usado o intervalo de confiança para proporções. O cálculo dos limites de um intervalo de confiança pode ser feito da seguinte forma (BUSSAB, 2012):

$$IC_{(1-\alpha)} = (\hat{p} - \epsilon; \hat{p} + \epsilon) \quad (15)$$

em que  $\hat{p}$  é a proporção estimada e  $\epsilon$  é o erro calculado na equação (16).

Considerando que 36 questionários respondidos é uma amostra pequena, usa-se o erro quando há correção para população finita. De acordo com Barbetta (2014), o cálculo se faz da seguinte forma:

$$\epsilon = Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \quad (16)$$

em que  $n$  é o tamanho da amostral,  $N$  é o tamanho da população finita,  $p$  é a proporção,  $Z$  é a variável aleatória normal padronizada,  $\alpha$  é o nível de confiança desejado e  $\epsilon$  é o erro padrão. Para todos os intervalos de confiança construídos utilizou-se 95% de confiança.

Dessa maneira, como há 60 professores atuando na Universidade Federal de Alfenas, *campus* Varginha, o erro em relação a população base é pequeno. Porém, os resultados serão todos com base na população, ou seja, não se pode generalizar os resultados para uma população que fuja muito do perfil desses professores.

Todas as perguntas do questionário foram feitas com base nas perguntas que as seguradoras brasileiras fazem a seus clientes antes de fechar um contrato de seguro. Para facilitar a visualização gráfica, as perguntas foram abreviadas em Perg1, Perg2 e assim por diante. As abreviações com as respectivas perguntas ficaram da seguinte forma:

- Perg1: Qual porcentagem que você pagaria a mais no seguro seu carro?
- Perg2: (Para quem não tem seguro) Qual porcentagem de desconto que o faria adquirir um seguro para seu carro?
- Perg3: Se o valor de mercado do seu carro fosse 50% maior do que o seu carro atual, qual a porcentagem que você pagaria a mais no seguro do carro?
- Perg4: Se o valor de mercado do seu carro fosse 100% maior do que o seu carro atual, qual a porcentagem que você pagaria a mais no seguro do carro?
- Perg5: Se o valor de mercado do seu carro fosse 200% maior do que o seu

carro atual, qual a porcentagem que você pagaria a mais no seguro do carro?

- Perg6: (Para quem tem garagem) Se sua residência não tivesse garagem, qual a porcentagem que você pagaria a mais no seguro do seu carro?
- Perg7: (Para quem reside na mesma cidade que trabalha) Se você não residisse na mesma cidade que trabalha, qual a porcentagem que você pagaria a mais no seguro do seu carro?
- Perg8: Qual a porcentagem que você pagaria a mais no seguro do seu carro, se o filho de 18 - 22 anos dirigisse com frequência o seu carro, ou seja, quase todo dia?
- Perg9: (Para quem é casado) Caso fosse solteiro, qual a porcentagem que você pagaria a mais no seguro do seu carro?
- Perg10: (Para quem é solteiro) Caso fosse casado, qual a porcentagem que você pagaria a mais no seguro do seu carro?

As variáveis escolhidas para a Análise de Componentes Principais foram todas as perguntas que envolvessem a ideia de risco além da variável Idade. As variáveis suplementares quantitativas escolhidas foram: Anos que trabalha na Unifal, Número de filhos, Se possui garagem, Se reside na mesma cidade em que trabalha e Se tinha

seguro na cidade em que morava antes, sendo que as quatro últimas são de resposta binária (0 para não, 1 para sim). As variáveis suplementares qualitativas escolhidas foram: sexo, estado civil, cidade, marca do carro e modelo do carro. Tais variáveis são apenas ilustrativas, não entram no cálculo dos componentes principais, apenas aparecem como forma de ilustração gráfica.

Para a análise gráfica foram utilizados os dois primeiros componentes principais, ou seja, aqueles que mais explicam o conjunto de dados originais. Como os componentes principais são todos ortogonais entre si, o gráfico com os dois primeiros componentes principais é mais fácil de analisar, pois representa os eixos das ordenadas e abcissas. Por se tratar de variáveis sensoriais, a Sensometria denomina os gráficos que projetam os dados nos dois primeiros componentes principais como Mapas de Preferências Internos. Além disso, foi usado cálculo de correlação de cada uma das variáveis originais, com os dois primeiros componentes principais, para averiguar qual delas mais explica esses componentes.

Todos os cálculos estatísticos e matemáticas foram feitos no programa gratuito R CORE TEAM (2016), além disso foi utilizado o pacote *FactoMineR* (HUSSON *et al*, 2015).

## 7. RESULTADOS E DISCUSSÃO

Após coletadas as respostas dos questionários, foi efetuada a Análise de Componentes Principais, de modo que fique fácil analisar as correlações entre as perguntas do questionário. Esses componentes foram construídos utilizando as perguntas que envolvessem a ideia de risco e a variável idade. Dessa maneira foi encontrada a matriz de autovalores, com base na matriz de variâncias e covariâncias das variáveis originais.

Os primeiros autovalores são, conforme visto na Revisão de Literatura, os que compreendem maior parte da variabilidade total das variáveis originais. Dessa maneira, é notável que o primeiro autovalor consegue explicar uma boa parte da variabilidade, pois seu valor foi calculado 5,848 e o segundo autovalor 1,327, quase cinco vezes menor, esses autovalores representam as variâncias dos dois primeiros componentes principais.

Os dois primeiros componentes principais foram responsáveis por explicar 65,23% variação total dos dados, o que, de certa forma, é uma porcentagem aceitável de explicação. O trabalho irá focar nos dois primeiros componentes principais,  $Y_1$  e  $Y_2$ , e os valores das correlações das variáveis com os dois primeiros componentes principais estão na Tabela 1.

**Tabela 1:** Correlações entre cada variável e os componentes 1 e 2.

Variável	1CP	2CP
Idade	-0,027	0,611
Perg1	0,925	-0,129
Perg2	-0,033	-0,535
Perg3	0,926	0,149
Perg4	0,776	0,414
Perg5	0,743	0,444
Perg6	0,818	-0,189
Perg7	0,912	-0,161
Perg8	0,88	-0,181
Perg9	0,585	-0,405
Perg10	0,601	0,046

**Fonte:** Autores.

De acordo com os valores da Tabela 1 é possível ver que as variáveis Idade e Perg2 foram as que menos contribuíram para construção do primeiro componente principal. Enquanto as outras variáveis ajudaram bastante na construção desse componente, sendo que a Perg1, Perg3 e Perg7 tiveram maior correlação com o primeiro componentes, assim, pode-se afirmar que as mesmas foram as mais significativas para essa componente. Ou seja, de certa forma essas perguntas, mais especificamente a Perg3 e a Perg7, foram as quais os professores responderam mais cautelosamente, sendo assim, os questionamentos sobre o valor do carro ser maior e também de não morar na mesma cidade resultaram em uma maior ideia de risco para os entrevistados.

A proporção de professores que pagariam 30% a mais em seus seguros de

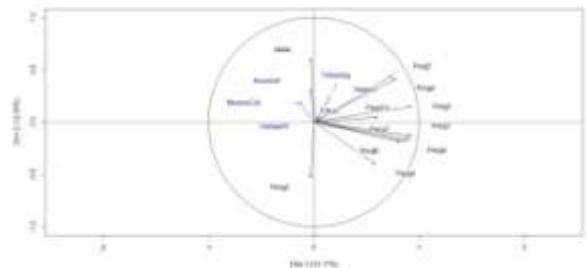
carro, caso o valor de seus carros fosse 50% maior (Perg3), é de, 36%, podendo estar entre 26% e 46% com 95% de confiança. As Perg4 e Perg5, que também estão relacionadas ao valor do carro, tiveram uma proporção alta de professores que pagariam 30% a mais em seus seguros, sendo 55,5% ( $\pm 10,35\%$ ) e 63% ( $\pm 10,05\%$ ) respectivamente. Assim, pode-se dizer que o valor do carro é um fator importante para os professores na percepção de risco dos mesmos e na perda financeira caso houvesse um sinistro. Já a proporção de professores que pagariam mais de 30% do valor do seguro do seu carro para as Perg9 e Perg10 foi de apenas 2,5%, admitindo estar entre 0,5% e 4,5%. Talvez essa baixa proporção explique a baixa correlação dessas variáveis com o primeiro componente.

Ainda pela Tabela 1, percebe-se que as variáveis Idade e Perg2 são muito mais importantes para construção do segundo componente principal que no primeiro. Porém, deve-se levar em consideração que a variável Perg2 teve apenas 2 observações, ou seja, apenas dois entrevistados responderam a essa pergunta, pois a maioria deles tem seguro de carro. Assim, apesar da Perg2 ajudar na construção do segundo componente principal, ela não é muito informativa para a construção dos componentes.

Como as perguntas que envolvem

a ideia de risco estão mais correlacionadas com o primeiro componente principal, os professores que foram mais avessos ao risco, ou seja, estariam dispostos a desembolsar mais dinheiro pelos seus seguros de carro contribuíram mais para a construção do primeiro do componente principal.

Graficamente, pode-se perceber as correlações entre as variáveis e os dois primeiros componentes principais analisando-se a Figura 1. Os vetores das perguntas estão todas apontando para direita, justamente porque as perguntas foram todas correlacionadas positivamente com o primeiro componente principal.



**Figura 1:** Mapa de Preferência Interno e os vetores das variáveis.

**Fonte:** Autores.

A Figura 1 apresenta os vetores das coordenadas (correlações) das variáveis originais projetadas nos dois primeiros componentes principais. O eixo horizontal é o primeiro componente principal e o eixo vertical é o segundo componente principal, o círculo de correlação vai de -1 a 1 (pois as correlações também tem esses limites). Quanto mais próximo o vetor estiver do limite do círculo, maior é a correlação (próximo de -

1 ou 1), de acordo com a direção que o vetor estiver apontando, para o primeiro componente (vetor na horizontal) ou para o segundo componente (vetor na vertical). Dessa forma, quanto mais próximo de 90° o ângulo entre algum vetor e um dos dois componentes, mais independente é a variável desse componente, ou seja, correlação pequena. Além disso, da mesma forma, ângulo pequeno entre os próprios vetores das variáveis significa alta correlação entre elas, e quanto maior o ângulo entre os vetores, menos correlacionados são os mesmos.

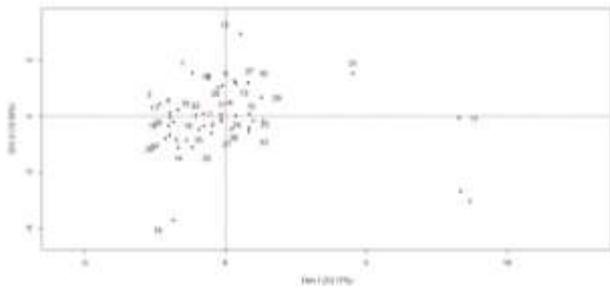
Analisando a Figura 1 e, tendo em vista que os componentes principais são ortogonais (independentes), verifica-se que a variável Idade é quase independente das outras variáveis, nesse caso as respostas dos professores da Universidade Federal de Alfenas *campus* Varginha, ou seja, suas percepções de risco não foram influenciadas pela idade.

Além disso, os professores tiveram um comportamento esperado, os entrevistados responderam de forma coerente os questionários, pois todas as perguntas estão relacionadas, ou seja, aqueles entrevistados que têm muita aversão ao risco, responderam mais receosos às perguntas, já os professores que não tem tanta aversão ao risco responderam de forma mais arrojada às situações impostas. Isso significa que entre os professores existem aqueles que são mais avessos ao risco e aqueles que são mais

corajosos diante dos riscos envolvendo seguro de carro.

Observando as variáveis suplementares quantitativas, na Figura 1, propostas para o trabalho, vê-se que elas também foram coerentemente respondidas. Os anos em que os professores trabalham na Universidade estão muito correlacionados à idade dos mesmos e, conseqüentemente, são indiferentes as respostas das perguntas. Além disso, é possível observar que os professores que têm garagem e moram na mesma cidade que trabalham, não são tão avessos ao risco, talvez por estarem mais protegidos contra esses riscos a percepção deles é afetada. Já os professores que têm seguro de carro, que já tinham seguro de carro na cidade onde moravam e que têm mais filhos estão dispostos a pagar mais pelo seguro de carro, o que faz todo sentido, pois os mesmos já têm seguro e tinham seguro antes, além disso quanto mais filhos, mais despesas, o que acarreta numa liberdade financeira menor, por isso eles se protegem mais de riscos que venham causar instabilidade econômica.

Também foram analisados os gráficos com as variáveis suplementares qualitativas: sexo, estado civil, estado de origem, marca do carro e tipo de carro. Elas foram colocadas individualmente com os gráficos de dispersão dos dados nos dois primeiros componentes principais. O mapa de preferência com a variável sexo está ilustrada na Figura 2.



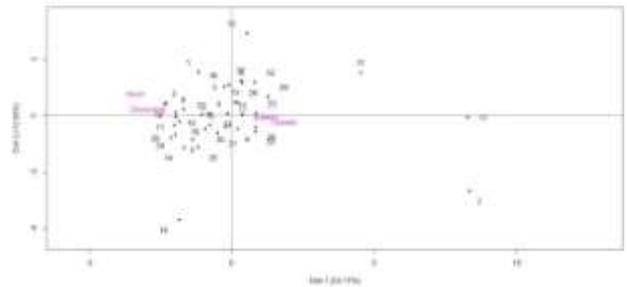
**Figura 2:** Mapa de Preferência Interno e a variável suplementar sexo.  
**Fonte:** Autores.

Sabendo que foram entrevistados 23 professores e 13 professoras, pode-se analisar a Figura 2 da seguinte forma: o baricentro dos dados relacionados ao sexo feminino está contribuindo mais para a construção do primeiro componente principal, por estar mais a direita, ou seja, pode-se dizer que as professoras da Universidade tendem a ser mais avessas ao risco do que os professores. Conforme Vallin (2004), isso não seria apenas uma especificidade dos professores da Universidade Federal de Alfenas, *campus* Varginha. Para ele, de forma geral, os homens correm mais riscos que as mulheres, isso pode ser explicado tanto por fatores genéticos, como nível de testosterona, quanto por fatores socioculturais.

Conforme Souza (2007), um seguro pode custar em média 30% a menos para as mulheres, nas mesmas condições que os homens. Com base no trabalho, pode-se cobrar das professoras um valor maior no seguro, pois elas pagam em média menos e estão mais dispostas a desembolsar dinheiro nos seguros de carro, conseguindo-se equilibrar o preço pago entre homens e

mulheres. Com isso, a seguradora não perde os clientes homens, que inicialmente pagariam um valor alto, e nem as mulheres.

Na Figura 3 está ilustrado o mapa de preferência interno com a variável estado civil.

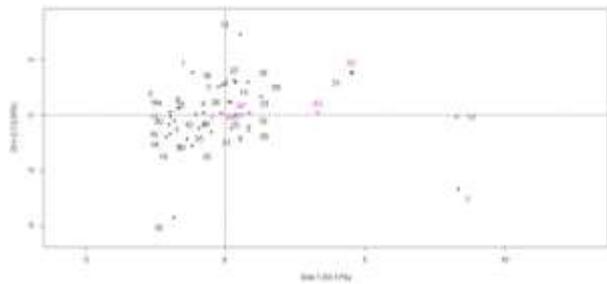


**Figura 3:** Mapa de Preferência Interno e a variável suplementar estado civil.  
**Fonte:** Autores.

Apenas um entrevistado colocou seu estado civil como Viúvo, dessa forma, não é possível tirar conclusões ou hipóteses sobre isso. Porém, pode ser visto que, de acordo com a Figura 3, os professores casados responderam às perguntas com mais aversão ao risco, já os professores divorciados têm mais preferência ao risco, ou seja, preferem correr risco do que gastar dinheiro com seguro de carro. Isso se deve ao fato de que os professores casados têm mais despesas, e por isso não queira assumir riscos prejudique suas finanças pessoais.

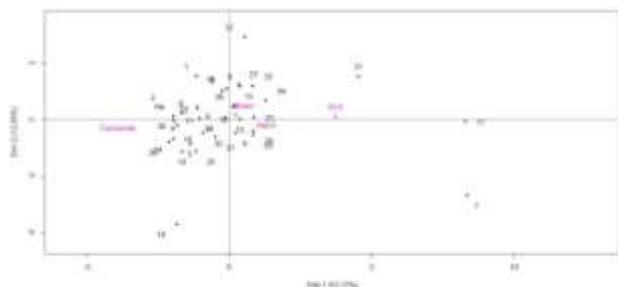
Souza (2007) mostra que as pessoas casadas têm, em média, um desconto de 15% em seus seguros, para as mesmas condições. Isso é exatamente o inverso do que pode ser visto no trabalho, os entrevistados casados tendem a querer se resguardar mais e, por isso, estão dispostos a pagar mais. Na Figura

4 tem-se o mapa de preferência interno juntamente com a variável suplementar estado de origem.



**Figura 4:** Mapa de Preferência Interno e a variável suplementar estado de origem.  
**Fonte:** Autores.

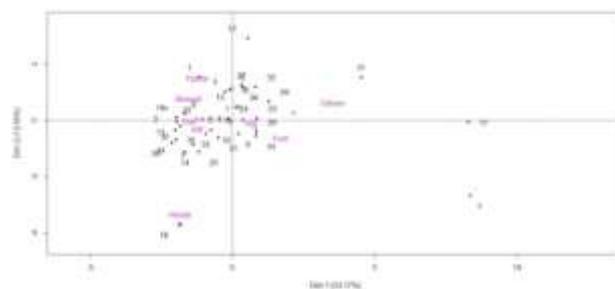
Tendo como base a Figura 4, pode-se afirmar que os professores que vieram do Rio de Janeiro são mais avessos ao risco. Obtiveram-se apenas dois entrevistados do Rio de Janeiro, ou seja, não se pode tirar uma conclusão concreta sobre a influência dos Estado na sua percepção de risco ligada aos automóveis. A Figura 5 mostra o mapa de preferência junto com a variável tipos de carro.



**Figura 5:** Mapa de Preferência Interno e a variável suplementar tipo de carro.  
**Fonte:** Autores.

Os tipos de carro que os professores possuem influenciam em suas percepções de risco, de acordo com a Figura 5. Os donos de *SUVs* e *Hatches* foram mais avessos ao risco, ou seja, responderam às perguntas dos

questionários de forma mais receosa. Talvez isso aconteça porque, de acordo com a Superintendência de Seguros Privados (2015), dos dez carros mais roubados no primeiro semestre de 2015, seis eram do tipo *Hatch*, dois *Sedan* e dois do tipo camionete. Isso pode ter feito os donos de carros *Hatch* serem mais cautelosos em relação aos riscos que eles podem estar correndo. Na Figura 6 é visto o mapa de preferência interno e a variável marca de carro.



**Figura 6:** Mapa de Preferência Interno e a variável suplementar marca do carro.  
**Fonte:** Autores.

Como poucos questionários foram respondidos em relação à diversidade de marcas de carro, 8 no total, fica difícil uma análise e uma conclusão mais concreta sobre a Figura 6, porém, pode ser dito que os professores donos de carros da marca *Citroen*, *Ford* e *GM* são, inicialmente, mais avessos ao risco e estariam dispostos a desembolsar mais dinheiro para se sentirem seguros em relação aos seus carros.

## 8. CONSIDERAÇÕES FINAIS

Com o trabalho foi possível analisar e quantificar as percepções que os

professores, da Universidade Federal e Alfenas *campus* Varginha, têm sobre os riscos que envolvem seguros carros. Foi possível analisar as diferentes aversões ao risco para diferentes perfis de consumidores. Por exemplo, é possível averiguar que os professores casados são mais avessos ao risco que os professores solteiros e divorciados.

Os resultados do trabalho podem ser úteis para seguradoras do ramo automobilístico, pois os mesmos podem deixar mais flexíveis as precificações. Por exemplo: uma seguradora que acaba cobrando mais das pessoas solteiras, de acordo com o modelo de precificação utilizado (modelos lineares generalizados, por exemplo), pode muito bem equilibrar o valor dos prêmios, ou seja, cobrar mais das pessoas casadas para não ter que cobrar tanto dos professores solteiros, porque, de acordo com os resultados, as pessoas casadas estão dispostas a gastar mais com seguro de carro. Dessa maneira, a seguradora acaba por não perder clientes solteiros devido ao preço do prêmio.

De certa forma, a seguradora poderá distribuir os riscos entre categorias de clientes, com base nos resultados desse trabalho. Feita a precificação, será possível avaliar quais grupos de pessoas tendem a sofrer mais sinistros e, conseqüentemente, pagar mais caro pelo seguro. Com isso, a seguradora transferiria os riscos desse grupo para um outro, que não tem tanta frequência

de sinistros, mas que estaria disposto a pagar um valor maior de prêmio.

Para trabalhos futuros, seria interessante aplicar esses questionários e as metodologias em uma população maior e averiguar se os resultados obtidos para os professores da Universidade Federal de Alfenas, *campus* Varginha, pode ser generalizados para uma população bem maior, como uma cidade ou estado. Além disso, seria ideia obter dados reais de seguradoras para tentar precificar, com base em um conjunto de apólices, os seguros para os usuários, utilizando essa flexibilidade que a Sensometria propõe.

## 9. REFERÊNCIAS BIBLIOGRÁFICAS

ANZALDÚA-MORALES, A. **La evaluación sensorial de los alimentos en la teoría y la práctica**. Editorial Acribia S. A., Zaragoza, 1994. 198 p.

BAILEY, R. A.; SIMON, L. J. Two studies in automobile insurance ratemaking. **Astin Bulletin**, v. 1, n. 4, p. 192-217, 1960.

BARBETTA, P. A. **Estatística Aplicada às Ciências Sociais**. 9 ed. Florianópolis: Editora UFSC, 2014. 320 p.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 7 ed. São Paulo: Editora Saraiva, 2012. 540 p.

CHANG, L.; FAIRLEY, W. B. Pricing Automobile Insurance under Multivariate Classification of Risks: Additive versus Multiplicative. **The Journal of Risk and Insurance**, v. 46, n. 1, p. 75-98, 1979.

DE JONG, P.; HELLER, G. Z. **Generalized linear models for insurance data.**

Cambridge: Cambridge University Press, 2008, 196 p.

DICKISON, D. C. M. **Insurance Risk and Ruin.** Cambridge: Cambridge University Press, 2005, 229 p.

DIONNE, G.; VANASSE, C. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. **Astin Bulletin**, v. 19, n. 2, p. 199-212, 1989.

FERREIRA, D. F. **Estatística Multivariada.** 2 ed. Lavras: UFLA, 2011. 676 p.

GARCINDO, L. A. M.; FERREIRA, E. B. Monitoramento e avaliação baseados em Sensometria para cursos de educação à distância. **Revista da Universidade do Vale do Rio Doce**, v.11, n.1, p. 465-490. 2013.

GRISTEC. **Dados de furto e roubos de veículos e cargas.** Disponível em: <[http://www.gristec.com.br/disco\\_virtual/Dados\\_Estatisticos.pdf](http://www.gristec.com.br/disco_virtual/Dados_Estatisticos.pdf)>, acessado em: novembro de 2014.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 6, p. 417-441, 1933.

HUSSON, F.; LÊ, S.; PAGÈS, J. **Exploratory Multivariate Analysis by Example Using R.** Chapman e Hall/CRC Press, 2011. 224 p.

HUSSON, F.; JOSSE, J.; LÊ, S.; MAZET, J (2015). **FactoMineR: Multivariate Exploratory Data Analysis and Data Mining.** R package version 1.31.3.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis.** 6 ed. New Jersey: Prentice Hall, 2007, 793 p.

MICHAELIS. **Dicionário de Português Online.** Melhoramentos, Disponível

em:<<http://michaelis.uol.com.br>>, acessado em: setembro de 2015.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, n. 6, p. 559-572, 1901.

PINDYCK, R. S.; RUBINFELD, D. L. **Microeconomia.** 6 ed. São Paulo: Pearson Prentice Hall, 2005, 641 p.

R Core Team (2016). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria.

REES, R.; WAMBACH, A. **The Microeconomics of Insurance.** Foundations and Trends in Microeconomics, vol. 4, 2008, 163 p.

SILVA, Y. R.; AFONSO, L. E. A Comparative Study of Pricing Methods of Automobile Insurance in Brazil. **Revista Brasileira de Risco e Seguro**, Rio de Janeiro, v. 11, n. 19, p. 25-44, 2015.

SOUZA, S. **Seguros: Contabilidade, Atuária e Auditoria.** São Paulo: Editora Saraiva, 2007, 229 p.

STORTI, L. B.; FERREIRA, E. B.; PEREIRA, C. A importância dos experimentos em faixas na Sensometria: o caso do queijo Minas Padrão com insulina. **Sigmae**, Alfenas, v. 3, n. 2, p. 25-33, 2014.

SUSEP. **IVR - Índice De Veículos Roubados.** Disponível em: <<http://www2.susep.gov.br/menuestatistica/RankRoubo/menu1.asp>>, acessado em: Outubro de 2015.

VALLIN, J.; PINELLI, A. Mortalidade, sexo e gênero. **Gênero nos estudos de população**, p. 15-54, 2004.

VARIAN, H. R. **Microeconomia: Uma abordagem moderna.** 8 ed. Rio de Janeiro: Elsevier, 2012, 821 p.