

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
Curso de Especialização em Estatística

**ESTUDO SOBRE A DISTRIBUIÇÃO ESPACIAL DE VALORES DE PRÊMIO DE  
SEGUROS DE VEÍCULOS NO TERRITÓRIO NACIONAL BRASILEIRO**

Poliane da Silva Duarte

**Belo Horizonte**  
**2011**

Poliane da Silva Duarte

**ESTUDO SOBRE A DISTRIBUIÇÃO ESPACIAL DE VALORES DE PRÊMIO DE  
SEGUROS DE VEÍCULOS NO TERRITÓRIO NACIONAL BRASILEIRO**

Monografia apresentada ao curso de Especialização em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística.

Orientador: Marcelo Azevedo Costa

**Belo Horizonte**

**2011**

**Poliane da Silva Duarte**

**ESTUDO SOBRE A DISTRIBUIÇÃO ESPACIAL DE VALORES DE PRÊMIO DE  
SEGUROS DE VEÍCULOS NO TERRITÓRIO NACIONAL BRASILEIRO**

Monografia apresentada ao curso de Especialização em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística.

.

---

Prof. Marcelo Azevedo Costa (Orientador) – UFMG

---

Prof. Ela Mercedes Medrano de Toscano – UFMG

---

Prof. Aureliano Angel Bressan – UFMG

**Belo Horizonte, 01 de outubro de 2011.**

Dedico este trabalho aos meus pais e irmãos, pelos ensinamentos de bons valores, amor incondicional, compreensão e incentivo, que me serviu de alicerce para alcançar mais este sonho.

## **AGRADECIMENTOS**

Ao Deus FIEL, por mais uma promessa que seu cumpriu em minha vida e que cuidou de cada detalhe para que o sonho se tornasse realidade.

Aos meus tesouros: pais e irmãos, pelo amor, dedicação e exemplos de caráter e integridade.

Ao Paulo Rocha dos Santos, meu futuro esposo, pelo incentivo, apoio e compreensão durante todo o período acadêmico.

Ao professor Marcelo Azevedo Costa, pela excelente orientação prestada para o desenvolvimento deste trabalho.

E aos meus queridos amigos pelo companheirismo, que mesmo na distância estiveram sempre presentes em todos os momentos da minha vida.

*“Podemos escolher recuar em direção à segurança ou avançar em direção ao crescimento. A opção pelo crescimento tem que ser feita repetidas vezes. E o medo tem que ser superado a cada momento”*

*Abraham Maslow*

## RESUMO

O presente trabalho trata-se de um estudo descritivo e analítico da disposição espacial dos valores de prêmio de seguros de veículos no território nacional brasileiro, baseando-se em dados disponibilizados pela Superintendência de Seguros Privados (SUSEP) associados à listagem de cep's nacionais brasileiros e suas respectivas coordenadas geográficas (latitude e longitude). Para fins deste estudo, foram também utilizados recursos do pacote estatístico R, além de técnicas específicas de varredura espacial de dados. Inicialmente foram feitas análises descritivas dos pontos para a verificação das características do conjunto, em seguida foi realizada a plotagem destes dados no mapa do Brasil a fim de se analisar a distribuição espacial dos pontos e finalmente, fez-se a análise da distribuição espacial destes dados por intensidade, a fim de verificar as regiões brasileiras onde se pode observar os maiores e menores valores de prêmio de seguros de automóvel. A fim de se realizar um estudo mais amplificado, tomou-se como exemplo automóveis de preços diversificados no mercado brasileiro.

**PALAVRAS-CHAVE:** Análise de agrupamento de dados. Estatística de varredura espacial. Distância Euclidiana. K-médias. Simulação de Monte Carlo. Verossimilhança. Seguro de veículos. Cálculo de prêmio. Exposição do risco.

## **ABSTRACT**

The present work it is a descriptive and analytical study of spatial arrangement of the values of premium vehicle insurance in Brazilian territory, based on data provided by the Superintendência de Seguros Privados (SUSEP) associated with the listing of zip's Brazilian nationals and their geographical coordinates (latitude and longitude). For purposes of this study were also used resources of the statistical package R, and specific techniques of spatial scan data. Initially descriptive analysis of the points made to verify the characteristics of the set, then the plot was carried out of the map data of Brazil in order to analyze the spatial distribution of points and finally made the analysis of the spatial distribution of these data by intensity, to verify the regions of Brazil where you can observe the highest and lowest values of auto insurance premium. In order to conduct a more amplified study, was taken as an example, car prices varied in the Brazilian market.

**KEYWORDS:** Cluster analysis of data. Spatial scan statistic. Euclidean distance. K-means. Monte Carlo simulation. Likelihood. Vehicle insurance. Calculation of prêmio. Risk exposure.

## LISTA DE FIGURAS

FIGURA 1 Distribuição pontual das observações da variável premio_exp no território nacional brasileiro.....	28
FIGURA 2 - Distribuição pontual das observações para os veículos Honda Civic no território nacional brasileiro.....	29
FIGURA 3 - Histogramas da variável premio_exp referente aos veículos Honda Civic antes e depois da associação das observações às coordenadas geográficas.....	30
FIGURA 4 - Distribuição pontual por intensidade dos valores de prêmio_exp dos veículos Honda Civic no território nacional brasileiro.....	33
FIGURA 5 - Comparação das suavizações com $k=25$ e $k=10$ da distribuição por intensidade dos valores de prêmio_exp dos veículos Honda Civic.....	35
FIGURA 6 - Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio_exp para os veículos Honda Civic no território nacional brasileiro.....	37
FIGURA 7 - Distribuição pontual das observações para os veículos Fiat Uno 1.0 no território nacional brasileiro.....	38
FIGURA 8 - Histogramas da variável premio_exp referente aos veículos Fiat Uno 1.0 antes e depois da associação das observações às coordenadas geográficas.....	39
FIGURA 9 - Distribuição pontual por intensidade dos valores de prêmio_exp dos veículos Fiat Uno 1.0 no território nacional brasileiro.....	42

FIGURA 10 - Comparação das suavizações com $k=25$ e $k=15$ da distribuição por intensidade dos valores de prêmio_exp dos veículos Fiat Uno 1.0.....	44
FIGURA 11 - Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio_exp para os veículos Fiat Uno 1.0 no território nacional brasileiro.....	46
FIGURA 12 - Distribuição pontual das observações para os veículos Fiat Siena 1.0 no território nacional brasileiro.....	47
FIGURA 13 - Histogramas da variável prêmio_exp referente aos veículos Fiat Siena 1.0 antes e depois da associação das observações às coordenadas geográficas.....	48
FIGURA 14 - Distribuição pontual por intensidade dos valores de prêmio_exp dos veículos Fiat Siena 1.0 no território nacional brasileiro.....	51
FIGURA 15 - Comparação das suavizações com $k=25$ e $k=10$ da distribuição por intensidade dos valores de prêmio_exp dos veículos Fiat Siena 1.0.....	53
FIGURA 16 - Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio_exp para os veículos Fiat Siena 1.0 no território nacional brasileiro.....	55

## LISTA DE TABELAS

TABELA 1 - Honda Civic: dados antes da associação à base de cep's.....	31
TABELA 2 - Honda Civic: dados após a associação à base de cep's.....	32
TABELA 3 - Fiat Uno 1.0: dados antes da associação à base de cep's.....	40
TABELA 4 - Fiat Uno 1.0: dados após a associação à base de cep's.....	41
TABELA 5 - Fiat Siena 1.0: dados antes da associação à base de cep's.....	49
TABELA 6 - Fiat Siena 1.0: dados após a associação à base de cep's.....	49

# SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Problema.....	14
1.2	Objetivo.....	14
1.2.1	<i>Objetivos Específicos</i> .....	14
1.3	Justificativa.....	15
2	REFERENCIAL TEÓRICO.....	16
2.1	Cálculo do prêmio em seguros de veículos.....	16
2.2	Exposição do risco.....	17
2.3	Análise de agrupamento de dados.....	18
2.3.1	<i>K-Médias</i> .....	19
2.3.2	<i>Distância Euclidiana</i> .....	20
2.4	Verossimilhança.....	20
2.5	Simulação de Monte Carlo.....	21
2.6	Estatística de varredura espacial para dados normais.....	21
2.6.1	<i>Cálculo da Verossimilhança</i> .....	22
3	METODOLOGIA.....	24
3.1	Tipo de pesquisa.....	25
4	ESTUDO DE CASO.....	26
4.1	Coleta e tratamento dos dados.....	26
4.2	Análise dos resultados.....	27
4.2.1	<i>Representação da disposição de todas as observações associadas da base no território nacional brasileiro</i> .....	28
4.2.2	<i>Análise dos dados referentes aos modelos Honda Civic</i> .....	28
4.2.2.1	Análise descritiva da variável premio_exp.....	29

4.2.2.2	Análise da distribuição da variável premio_exp por intensidade .....	32
4.2.2.3	Distribuição dos pontos por aproximação .....	34
4.2.2.4	Aplicação da Estatística de Varredura Espacial.....	37
4.2.3	<i>Análise dos dados referentes aos modelos Fiat Uno 1.0.....</i>	<i>38</i>
4.2.3.1	Análise descritiva da variável premio-exp .....	39
4.2.3.2	Análise da distribuição variável premio_exp por intensidade .....	41
4.2.3.3	Distribuição dos pontos por aproximação .....	43
4.2.3.4	Aplicação da Estatística de Varredura Espacial.....	46
4.2.4	<i>Análise dos dados referentes aos modelos Fiat Siena 1.0 .....</i>	<i>47</i>
4.2.4.1	Análise descritiva da variável premio_exp .....	48
4.2.4.2	Análise da distribuição da variável premio_exp por intensidade .....	50
4.2.4.3	Distribuição dos pontos por aproximação .....	52
4.2.4.4	Aplicação da Estatística de Varredura Espacial.....	54
5	CONCLUSÃO .....	56
	REFERÊNCIAS .....	57
	ANEXO .....	59

## 1 INTRODUÇÃO

A precificação dos seguros de automóveis é uma atividade que exige a análise de vários critérios inerentes ao veículo em questão. Variáveis como categoria do veículo, modelo, ano de fabricação, preço de mercado, região de localização, perfil do condutor e índice de sinistralidade são imprescindíveis na determinação do valor do prêmio a ser cobrado pelo seguro de um determinado automóvel.

Utilizando-se do resultado do estudo destas variáveis e de outras como, por exemplo, taxa de carregamento, despesas administrativas e comerciais e custo de apólice, as seguradoras definem o preço, ou melhor, tecnicamente dizendo, determinam o prêmio a ser cobrado pelo seguro de um dado veículo.

Para a definição do valor do prêmio, utiliza-se tanto de técnicas atuariais quanto estatísticas, além de ser também necessária a observância de regras pré-estabelecidas pelo órgão competente SUSEP (Superintendência de Seguros Privados), que tem a função de fiscalizar e controlar o mercado de seguros no Brasil, entre outras atividades correlatas.

Semestralmente, todas as seguradoras brasileiras devem repassar o resultado das variáveis das suas carteiras, de acordo com características pré-definidas, para esta superintendência, que os divulga para consulta pública.

Por meio das informações publicadas referentes ao primeiro semestre do ano de 2010, no presente trabalho foram feitas análises dos valores dos prêmios com base no período de exposição do veículo ao risco, de acordo com a região de localização do automóvel. Após esta análise, foram realizadas observações acerca da variação dos valores destes prêmios no âmbito regional.

Especificamente para fins deste estudo, foram considerados apenas os automóveis referentes aos modelos Honda Civic, Fiat Uno 1.0 e Fiat Siena 1.0 fabricados a partir do ano de 2010, com o intuito de obter resultados para veículos novos e de custos diversificados no mercado.

## 1.1 Problema

Investigar a distribuição espacial dos valores de prêmios de seguros de automóveis, procurando verificar a existência de *clusters* nas localidades que apresentam os maiores ou menores valores de prêmios dentro do território nacional brasileiro. Neste sentido, coloca-se em questão: Existem conglomerados espacialmente localizados nas regiões que, comparadas às demais localidades do território nacional brasileiro, apresentam os maiores ou menores valores de prêmios de seguros de automóveis?

## 1.2 Objetivo

Desenvolver uma metodologia para visualização e análise da existência de conglomerados espaciais dos valores de prêmios de seguros de automóveis, no território nacional brasileiro, para os seguintes modelos de veículos: Honda Civic, Fiat Uno 1.0 e Fiat Siena 1.0.

### 1.2.1 Objetivos Específicos

- ✓ Analisar a distribuição dos valores de prêmio de seguros de veículos no território nacional brasileiro;
- ✓ Utilizar o pacote RGoogleMaps do software R para a visualização dos dados de seguros de veículos disponibilizados pela SUSEP no ano de 2010;

- ✓ Utilizar o método *Scan* de varredura espacial para delinear conglomerados para os elevados valores e conglomerados para os baixos valores de prêmios dos seguros de veículos no Brasil.

### 1.3 Justificativa

O interesse por informações referentes ao comportamento dos preços de mercado dos bens de consumo é característico à população de todas as classes sociais de acordo com as suas respectivas necessidades. Saber onde encontrar os menores preços de um determinado produto ou serviço é um dos maiores objetivos de todo consumidor e, sabe-se que devido à competitividade, quanto maior é a demanda por um determinado bem, maior torna-se também a variação do seu preço no mercado.

Desta forma, o automóvel que atualmente trata-se de um bem de grande procura dos brasileiros, também se tornou alvo das pesquisas de preços e, por ser um produto de valor consideravelmente elevado, a preocupação com a aquisição de um seguro que garanta proteção contra riscos que podem ocasionar a perda ou dano deste bem, também já é objeto de igual importância nestas pesquisas.

Sendo assim, diante da inexistência de estudos que avaliem o comportamento espacial de valores de seguros de veículos no território nacional brasileiro, o presente trabalho dedicou-se à esta finalidade, procurando definir em quais regiões do país estão localizados os maiores e menores preços para a prestação deste serviço.

## 2 REFERENCIAL TEÓRICO

### 2.1 Cálculo do prêmio em seguros de veículos

O prêmio corresponde ao valor referente ao custo do seguro.

De acordo com Ferreira (2002) prêmio é a prestação pecuniária, ou remuneração, que o segurado paga ao segurador para que este assuma o risco e pague a indenização no caso da ocorrência do evento segurado.

Os seguros são classificados em ramos de acordo com o bem a ser coberto e para cada ramo são consideradas variáveis específicas no cálculo do prêmio, conforme as características inerentes a este bem.

Segue abaixo a equação de cálculo do prêmio em seguros de veículos, conforme Ferreira (2002):

$$P_T = \left[ \frac{E[S] + DS - SALV + \theta}{1 - \alpha} + \delta^* \right] (1 + IOF)$$

Onde:

$P_T$  : Prêmio total

$E[S]$  : Esperança de sinistro (média de sinistros)

$DS$  : Despesa com sinistro

$SALV$  : Salvados e ressarcimentos

$\theta$  : Margem de segurança para proteção contra eventuais oscilações em torno dos sinistros esperados.

$\delta$  : Carregamento absoluto

$\delta^*$  : Custo de apólice

$\alpha$  : Soma das despesas administrativas, comissão de corretagem e margem de lucro do segurador.

*IOF*: Imposto sobre Operações Financeiras

Com relação às partes integrantes do cálculo do prêmio total em seguros de veículos, Ferreira (2002) afirma que as despesas com sinistro referem-se às inerentes à regulação (análise e definição do processo e sinistro), despesas de inspeção (vistorias e análise do risco sinistrado), honorários judiciais, peritagens, etc. Os salvados e ressarcimentos constituem, respectivamente, os bens restantes após a liquidação do sinistro e as devoluções referentes às vistorias judiciais contra terceiros que foram, comprovadamente, definidos culpados. Já o carregamento absoluto constitui-se pelas despesas de marketing com lançamento de produtos, custo de vistoria prévia, custo de atendimento em call center, etc.

## 2.2 Exposição do risco

Exposição do risco é o tempo em que um risco ficou exposto no período de análise. O valor atribuído a esta variável será sempre menor ou igual a 1, pois o cálculo aplicado consiste na razão entre o número de dias em interseção com o período de análise considerado no estudo e o número total de dias deste período.

Segue abaixo a equação de cálculo desta variável:

$$Exp = \frac{N^{\circ} \text{ de dias em interseção com o período de análise}}{N^{\circ} \text{ de dias do período de análise}}$$

Nos estudos atuariais sobre seguros, o período de análise geralmente considerado corresponde a 365 dias contados entre o primeiro e último dia do ano em questão.

### 2.3 Análise de agrupamento de dados

A análise de agrupamento consiste em uma das técnicas existentes para estudos estatísticos de dados. De acordo com Mingoti (2005), a análise de agrupamento, que é também conhecida como análise de conglomerados, classificação ou *Cluster*, consiste na segregação do conjunto de dados analisado em grupos, de acordo com características semelhantes dos elementos entre si, de forma que cada grupo gerado distingue-se dos demais com base nas variáveis que os separaram.

A divisão dos grupos na análise de agrupamento pode ser feita de várias formas diferentes. Segundo Mingoti (2005), as medidas mais comuns para dados quantitativos são: Distância Euclidiana, Distância Generalizada ou Ponderada, Distância de Minkowsky, Coeficiente de Concordância Simples, Coeficiente de Concordância Positiva, Coeficiente de Concordância de Jaccard e Distância Euclidiana Média. Todas estas medidas consistem em técnicas de dissimilaridade, desta forma, os dados comparados serão mais similares quanto menor for o valor encontrado no resultado da medida aplicada.

Além das medidas de dissimilaridade, são também utilizadas na análise de agrupamento técnicas aglomerativas que estão relacionadas ao processo de agrupamento dos dados, pois

(...) no início do processo de agrupamento tem-se n conglomerados, ou seja, cada elemento do conjunto de dados observado é considerado como sendo um conglomerado isolado. Em cada passo do algoritmo, os elementos amostrais vão sendo agrupados, formando novos conglomerados até o momento no qual todos os elementos considerados estão num único grupo. (...) Em cada estágio do procedimento de agrupamento, os grupos são comparados através de alguma medida de similaridade ( ou dissimilaridade) previamente definida. (MINGOTI, 2005, p.165)

As técnicas aglomerativas subdividem-se em hierárquicas e não hierárquicas. De acordo com Mingoti (2005), nas técnicas hierárquicas, elementos unidos num mesmo *cluster* durante o processo de agrupamento não poderão ser separados. Constituem estas técnicas os Métodos de Ligação Simples, Completa e Média, o Método do Centróide e o de Ward. Quanto as técnicas não hierárquicas, tratam-se de métodos nos quais o número de *clusters* é previamente definido, conforme o desejo do analista. Além disto, esta técnica permite a análise de conjunto de dados mais robustos, o que não é possível nos métodos hierárquicos. Dentre os tipos de técnicas não hierárquicas incluem-se o método das K-Médias, *Fuzzy c-Médias* e redes neurais artificiais.

### **2.3.1 K-Médias**

O método de agrupamento K-médias é um dos mais aplicados para agrupamento de dados. De acordo com Mingoti (2005), consiste na alocação de cada elemento amostral ao *cluster* que possui o seu vetor de média amostral mais próximo do vetor dos valores observados.

A formação dos grupos é feita através da junção das características dos elementos mais próximos. Desta forma, se para um determinado elemento não sabemos quais são as suas verdadeiras características, este assumirá as mesmas do dado que estiver mais próximo para os qual se pode identificar os atributos.

Originalmente, o método é composto por quatro passos:

- 1) Primeiramente escolhe-se  $k$  centróides, chamados de “sementes” ou “protótipos”, para se inicializar o processo de partição;
- 2) Cada elemento do conjunto de dados é, então, comparado com cada centróide inicial, através de uma medida de distância que, em geral, é a distância Euclidiana. O elemento é alocado ao grupo cuja distância é a menor.
- 3) Depois de aplicar o passo 2 para cada um dos  $n$  elementos amostrais, recalcula-se os valores dos centróides para cada novo grupo formado, e repete-se o passo 2, considerando-se os centróides destes novos grupos.

- 4) Os passos 2 e 3 devem ser repetidos até que todos os elementos amostrais estejam “bem alocados” em seus grupos, isto é, até que nenhuma realocação de elementos seja necessária.

(MINGOTI, 2005, p.165)

É importante colocar que existem várias formas distintas de implementação da técnica das k-médias. Conforme Mingoti (2005), o resultado final do agrupamento é influenciado pela seleção das sementes iniciais de agrupamento. Desta forma, deve-se atentar no momento desta escolha.

### 2.3.2 Distância Euclidiana

A distância Euclidiana, geralmente utilizada na aplicação do método das k-médias para comparação dos elementos, é definida como se segue, de acordo com Mingoti (2005):

Distância Euclidiana entre dois elementos  $X_L$  e  $X_k$ ,  $L \neq k$ , é definida por:

$$d(X_L, X_k) = [ (X_L - X_k)' (X_L - X_k) ]^{1/2} = [ \sum_{i=1}^p (X_{iL} - X_{ik})^2 ]^{1/2}$$

Ou seja, os dois elementos são comparados em cada variável pertencente ao vetor de observções.

## 2.4 Verossimilhança

De acordo com Montgomery e Runger (2009) a verossimilhança é a probabilidade de ocorrência dos valores da amostra.

O estimador da máxima verossimilhança é o que maximiza, em relação aos parâmetros da distribuição, a probabilidade de se obter os valores da amostra, segundo Montgomery e Runger (2009).

Mais adiante segue a equação para o cálculo desta estatística, dentro dos tópicos tratados nesta pesquisa.

## 2.5 Simulação de Monte Carlo

A simulação de Monte Carlo consiste em um método de geração de observações aleatórias de uma determinada variável com base em uma distribuição de probabilidade definida, afirmam Moore e Weatherford (2005).

De acordo com Kulldorff, Huang e Konty (2009), além do embasamento em uma distribuição de probabilidade pré-definida, a geração de observações aleatórias pela simulação de Monte Carlo também pode ser feita através da permutação dos valores de uma determinada amostra da variável em estudo.

## 2.6 Estatística de varredura espacial para dados normais

Kulldorff, Huang e Konty (2009) afirmam que o método estatístico de varredura é aplicado para análise de dados contínuos, considerando-se a localização espacial destes dados, com a observação das coordenadas de latitude e longitude, sendo que cada localização apresenta uma ou mais observações.

$x_s = \sum_{i \in S} x_i$ : número de de valores observados

$n_s$ : número de observações na localização

$x = \sum_i x_i$ : Soma de todos os valores observados

A representação dos grupos no espaço é feita através de círculos, para os quais é calculada uma razão do log da verossimilhança e o teste estatístico é definido como a máxima razão do log da verossimilhança sobre todos os círculos, afirmam Kulldorff, Huang e Konty (2009).

Os círculos estão centrados em uma observação e com um raio variando continuamente de zero até um determinado limite superior. Segundo Kulldorff, Huang e Konty (2009), para assegurar que serão encontrados tanto grupos pequenos quanto grandes, o limite superior é definido normalmente de forma que um círculo contenha no máximo 50% (cinquenta por cento) das observações. Este percentual somente é ultrapassado se houver um número grande de observações com características semelhantes, então o número de elementos divergentes para formação do outro grupo será menor. Vale ressaltar que círculos com apenas uma observação são ignorados.

O tamanho dos círculos é definido através de unidades específicas de distância como quilômetros (KM), por exemplo.

### 2.6.1 Cálculo da Verossimilhança

Conforme Kulldorff, Huang e Konty (2009), sobre o hipótese nula, a estimativa da máxima verossimilhança da média e variância são  $\mu = X/N$  e  $\sigma^2 = \frac{\sum_i (\mu - x_i)^2}{N}$ , respectivamente.

A verossimilhança sobre a hipótese nula é então:

$$L_0 = \prod_i \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

E o log da verossimilhança é:

$$\ln L_0 = -N \ln(\sqrt{2\pi}) - N \ln(\sigma) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$$

De acordo com Kulldorff, Huang e Konty (2009), sobre a hipótese alternativa, calcula-se os estimadores de máxima verossimilhança que são específicos para cada círculo  $z$ . A estimativa da máxima verossimilhança para a variância comum é:

$$\sigma_z^2 = \frac{1}{N} \left( \sum_{i \in z} x_i^2 - 2x_z \mu_z + n_z \mu_z^2 + \sum_{i \notin z} x_i^2 - 2(X - x_z) \lambda_z + (N - n_z) \lambda_z^2 \right)$$

Onde:

$$\mu_z = \frac{x_z}{n_z} : \text{m\u00e9dia dentro do c\u00edrculo}$$

$$\lambda_z = \frac{(X - x_z)}{(N - n_z)} : \text{m\u00e9dia fora do c\u00edrculo}$$

O log da verossimilhan\u00e7a para o c\u00edrculo z \u00e9:

$$\ln L(z) = -N \ln(\sqrt{2\pi}) - N \ln \left( \sqrt{\sigma_z^2} \right) - N/2$$

Como teste estat\u00edstico usa-se a raz\u00e3o da m\u00e1xima verossimilhan\u00e7a

$$\max_z \left( \frac{L_z}{L_0} \right)$$

Ou mais convenientemente, a raz\u00e3o do log da m\u00e1xima verossimilhan\u00e7a:

$$\max_z \left( \frac{\ln L_z}{\ln L_0} \right)$$

$$= \max_z \left( N \ln(\sqrt{2\pi}) - N \ln \left( \sqrt{\sigma_z^2} \right) - N/2 + N \ln(\sqrt{2\pi}) + N \ln \left( \sqrt{\sigma_0^2} \right) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2} = \right)$$

$$= \max_z \left( N \ln \left( \frac{\sigma_0}{\sigma_z} \right) + \sum_i \frac{(x_i - \mu)^2}{2\sigma^2} - \frac{N}{2} - N \ln \left( \sqrt{\sigma_z^2} \right) \right)$$

Atrav\u00e9s desta f\u00f3rmula, pode-se verificar que o *cluster* de maior verossimilhan\u00e7a selecionado \u00e9 o que minimiza a vari\u00e2ncia sobre a hip\u00f3tese alternativa.

Kulldorff, Huang e Konty (2009) afirmam que a signific\u00e2ncia estat\u00edstica do *cluster* de maior verossimilhan\u00e7a \u00e9 avaliada usando-se o teste de hip\u00f3tese de Monte Carlo. Ao inv\u00e9s de gerar dados aleat\u00f3rios da distribui\u00e7\u00e3o normal, um grande conjunto de dados

aleatórios é criado por permutação dos valores observados e suas respectivas localizações. Devido a esta aleatorização, o nível alfa correto será mantido mesmo se a observação não seguir a distribuição normal de fato.

Para cada conjunto de dados aleatório, é calculada a razão da log verossimilhança, a fim de verificar o *cluster* mais verossímil, afirmam Kulldorff, Huang e Konty (2009), que também apontam que, se a razão da log verossimilhança do conjunto de dados reais estiver entre as 5% maiores de todos os conjuntos, então o *cluster* de maior verossimilhança do conjunto de dados reais é estatisticamente significativo ao nível alfa de 5%.

Em suma, a estatística de varredura para dados normais busca identificar conjuntos de valores excepcionalmente altos, bem como os valores excepcionalmente baixos. Ainda que os dados não sigam a distribuição normal, o que faz com que não se possa mais obter um teste de razão de verossimilhança, o nível alfa correto é mantido, por isso, esta técnica pode ser aplicada a uma ampla variedade de dados contínuos, embora não seja indicada para dados exponenciais e outros tipos de dados de sobrevivência, para os quais existem métodos de varredura estatística apropriada.

A técnica estatística de varredura normal encontra-se livremente disponível no *software* SaTSacn, versão 7.0.

### **3 METODOLOGIA**

Neste capítulo será especificada a metodologia aplicada no presente estudo, a qual pode ser definida como o conjunto dos métodos a serem utilizados na construção de um trabalho de pesquisa. A metodologia busca apresentar de forma sistemática o universo de pesquisa, tornando a compreensão do estudo realizado mais acessível ao leitor.

Segundo Marconi e Lakatos (2003, p.83) o método consiste no “conjunto de atividades sistemáticas e racionais que, com maior segurança e economia, permite

alcançar o objetivo – conhecimentos válidos e verdadeiros-, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista.”

### **3.1 Tipo de pesquisa**

Segundo Vergara (2003) uma pesquisa pode ser classificada quanto aos seus objetivos e procedimentos de coleta de dados.

Quanto aos objetivos, este estudo é classificado como descritivo. De acordo com Vergara (2003), uma pesquisa descritiva tem como principal objetivo analisar com profundidade um determinado fenômeno, situação, procedimento ou fato ocorrido.

Com base na definição proposta pelo autor, esta classificação pode ser atribuída, pois esta pesquisa dedicou-se à análise detalhada da distribuição espacial de valores de prêmio de seguros de veículos no território nacional brasileiro a partir do estudo de uma base de dados específica.

Com relação aos procedimentos, este trabalho é qualificado como estudo de caso que é “o circunscrito a uma ou poucas unidades, entendidas essas como pessoa, família, produto, empresa, órgão público, comunidade ou mesmo país. Tem caráter de profundidade e detalhamento. Pode ou não ser realizado no campo.” (VERGARA, 2003, p.49)

Esta qualificação é conferida ao presente trabalho, visto que tem como foco o custo do seguro de automóveis no território brasileiro.

### **3.2 Universo e amostra**

Determinar o universo de pesquisa significa identificar qual será a população de estudo, levando-se em consideração suas características semelhantes, tais como sexo, faixa etária, estrutura de inserção, etc., conforme define Marconi e Lakatos (2003).

Para fins da realização deste trabalho tomar-se-á como universo de pesquisa, todo o território nacional brasileiro.

“População amostral ou amostra é uma parte do universo (população) escolhida segundo algum critério de representatividade.” (VERGARA, 2003, p.50)

A amostra dos dados necessária a este estudo refere-se ao primeiro semestre do ano de 2010, sendo composta pelo resultado da carteira de automóveis das seguradoras brasileiras.

## **4 ESTUDO DE CASO**

### **4.1 Coleta e tratamento dos dados**

Os dados utilizados neste trabalho compõem o resultado do primeiro semestre do ano de 2010 relativo aos seguros de veículos das apólices vigentes e sinistros ocorridos nas seguradoras brasileiras. Tais informações foram retiradas do site (<http://www.susep.gov.br/principal.asp>) da SUSEP, no link downloads, base Autoseg atualizada até o 1º semestre do ano de 2010. Trata-se de um banco de dados em Access, cujo nome é *BaseAuto*, do qual foi utilizada apenas a tabela *Arq\_Casco 3*.

Para fins deste estudo, empregaram-se apenas os resultados das variáveis exposição e prêmio, classificadas de acordo com a categoria, modelo, ano, região, cidade e CEP de circulação do veículo, sendo os códigos de modelos referentes à codificação padronizada da tabela FIPE.

Com intuito de se estudar a distribuição dos valores de prêmio por exposição, foi criada a variável *premio\_exp*, que consiste na razão entre as médias das variáveis prêmio e exposição.

Especificamente neste trabalho, foram estudados apenas os modelos Honda Civic, Fiat Uno 1.0 e Fiat Siena 1.0 fabricados a partir do ano de 2010, constando na base, respectivamente, 9.948, 36.551 e 18.741 veículos.

A fim de plotar os dados no mapa do Brasil, utilizou-se também de uma base composta pela listagem de diversos cep's do país e suas respectivas coordenadas geográficas (latitude e longitude), obtida através do site [www.mbi.com.br](http://www.mbi.com.br). Neste intuito, foi realizada a associação de todos os CEP's presentes na tabela Arq\_Casco 3 com esta base de coordenadas, ao que foram obtidos, respectivamente, apenas 1.326, 4.031 e 2.529 pontos associados do total de 9.948, 36.551 e 18.741 presentes na tabela da Susep.

Apesar da significativa perda de informações, o número de pontos combinados foi suficiente para mostrar a variação dos valores de prêmio por exposição dos seguros de veículos nas regiões brasileiras. Desta forma, este estudo baseou-se apenas nos conjuntos dos dados associados.

Ao plotar os dados no mapa do Brasil, foi verificado que os pontos associados dos três modelos em estudo não estão bem distribuídos no território brasileiro. Desta forma, com o intuito de se obter os valores de prêmio de seguros de veículos em todo o Brasil, foram gerados resultados por aproximação dos pontos existentes. Para tanto, aplicou-se a técnica de agrupamento não hierárquica feita pelo método das K-Médias, que utiliza a medida de Distância Euclidiana para comparar os elementos da amostra.

Por fim, na intenção de verificar as regiões brasileiras onde se pode observar os maiores e menores valores de prêmio de seguros de automóvel, os dados associados foram submetidos à técnica estatística de varredura espacial para definição dos conglomerados.

## **4.2 Análise dos resultados**

Nesta subseção serão apresentadas as análises descritivas dos dados e as observações feitas sobre os resultados gráficos obtidos com a utilização do pacote RgoogleMaps e do método estatístico de varredura espacial.

#### 4.2.1 Representação da disposição de todas as observações associadas da base no território nacional brasileiro



**Figura 1: Distribuição pontual das observações da variável premio\_exp no território nacional brasileiro**

Fonte: Elaborada pela autora

A figura 1 representa a disposição dos dados associados da base em estudo no território nacional brasileiro. Os pontos azuis compreendem a localização das observações da variável premio\_exp para todos os veículos constantes na base da SUSEP para os quais encontrou-se as coordenadas geográficas, ou seja, todos os pontos associados dos diversos modelos de veículos da base em questão. Verifica-se a predominância nas regiões Nordeste, Sudeste, Sul e relativamente no Centro-Oeste.

#### 4.2.2 Análise dos dados referentes aos modelos Honda Civic



**Figura 2: Distribuição pontual das observações para os veículos Honda Civic no território nacional brasileiro**

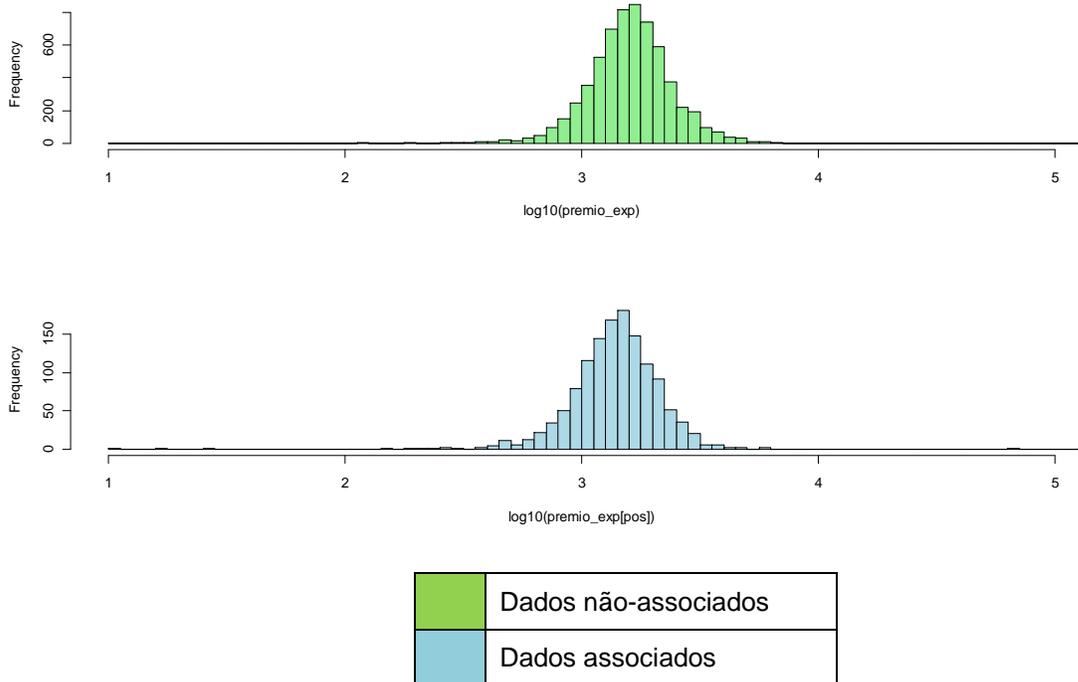
**Fonte: Elaborada pela autora**

A figura 2 se trata da representação da distribuição dos veículos relativos aos modelos Honda Civic. A disposição dos pontos indica que há uma maior concentração destes veículos nas regiões Sul e Sudeste do país.

#### **4.2.2.1 Análise descritiva da variável prêmio\_exp**

Conforme mencionado anteriormente, para fins deste estudo foram trabalhadas apenas as variáveis prêmio e exposição, mais especificamente a razão entre a primeira e a segunda, obtendo-se desta forma, a variável prêmio\_exp. Isso porque a intenção é a de se estudar o custo do seguro conforme o período de exposição do veículo ao risco.

Para uma melhor representação dos dados, trabalhou-se com o logaritmo de base dez ( $\log_{10}$ ) dos dados. Abaixo segue a análise desta variável:



**Figura 3: Histogramas da variável premio\_exp referente aos veículos Honda Civic antes e depois da associação das observações às coordenadas geográficas**  
**Fonte: Elaborada pela autora**

Comparando-se os histogramas acima, representantes dos valores de premio\_exp nos modelos Honda Civic, respectivamente, antes e depois do vínculo dos dados às coordenadas geográficas, verifica-se que os mesmos são bastante semelhantes, ou seja, o histograma que representa todos os dados da base, aproxima-se do que apresenta apenas os dados para os quais foram localizadas as coordenadas, o que indica que o número de dados associados, ou seja, a amostra obtida é satisfatória para o estudo, uma vez que sua representação aproxima-se do conjunto total dos dados.

## Análise descritiva dos dados antes da associação à base de cep's

**TABELA 1**  
Honda Civic: dados antes da associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
4,20	1.236,00	1.596,00	1.849,00	2.043,00	570.100,00	26,00	0,935	2,2e-16

Fonte: Elaborada pela autora

Conforme a figura 3 verifica-se que o histograma da variável premio\_exp referente aos dados não-associados apresenta-se relativamente simétrico.

Através das estatísticas encontradas, observa-se que os valores desta variável giram em torno da média de R\$ 1.849,00 e que 50% deles situam-se acima de R\$ 1.596,00 e 50% abaixo.

Observa-se que o menor valor encontrado para esta variável é R\$ 4,2, sendo que 25% estão dispostos abaixo de R\$ 1.236,00 e 25% estão acima de R\$ 2.043,00, sendo o valor máximo igual a R\$ 570.100,00.

A verificação de um valor mínimo significativamente baixo (R\$ 4,2) e valor máximo notavelmente alto (R\$ 570.100,00) pode estar relacionada ao curto período de exposição analisado, que influencia na razão prêmio/exposição, ou a um possível erro de digitação de valores na base de dados em estudo.

O teste de normalidade aplicado indica que a variável premio\_exp referente aos dados não-associados não segue a distribuição normal, o que se pode verificar mediante ao baixo valor da estatística p-value, inferior a 0,05.

## Análise descritiva dos dados após a associação à base de cep's

**TABELA 2**  
Honda Civic: dados após a associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
10,74	1.261,00	1.600,00	1.788,00	2.017,00	136.100,00	13,00	0,9049	2,2e-16

Fonte: Elaborada pela autora

Através da figura 3, o histograma da variável premio\_exp referente aos dados associados também se apresenta relativamente simétrico. As estatísticas encontradas indicam que os valores desta variável giram em torno da média de R\$ 1.788,00, estando 50% deles dispostos acima de R\$ 1.600,00 e 50% abaixo. O valor mínimo encontrado para esta variável é R\$ 10,74.

Observa-se que 25% dos dados estão abaixo de R\$ 1.261,00 e 25% estão acima de R\$ 2.017,00, sendo o valor máximo encontrado igual a R\$ 136.100,00.

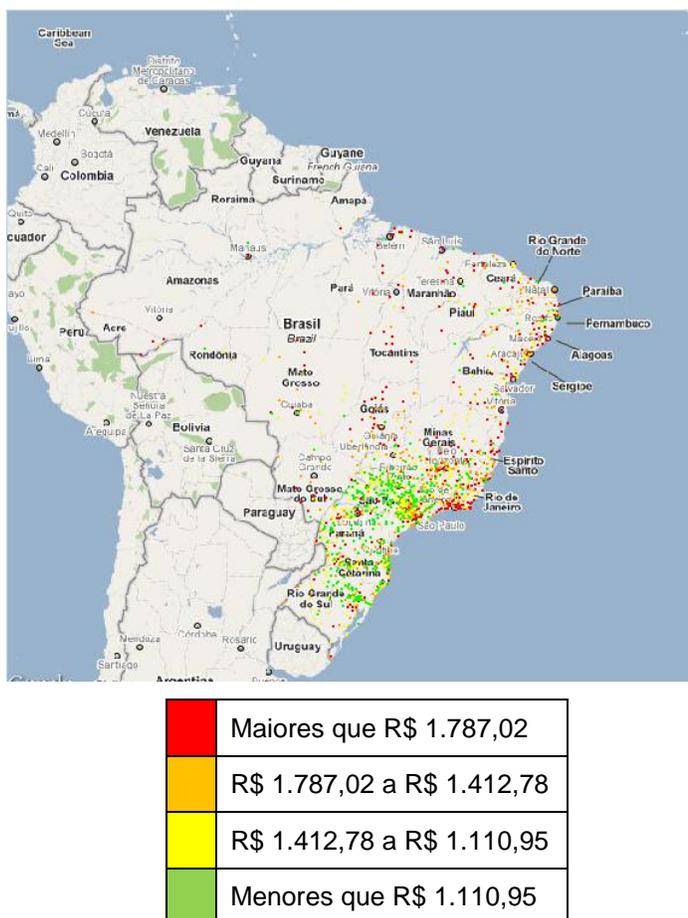
O resultado de um valor mínimo significativamente reduzido e de um valor máximo notoriamente elevado, justifica-se pelas mesmas razões apontadas na análise feita antes da associação à base de cep's.

O teste de normalidade aplicado indica que a variável premio por exposição referente aos dados associados não segue a distribuição normal, apresentando p-value inferior a 0,05.

Pode ser observado que a diferença no resultado da maior parte das estatísticas encontradas em ambas as análises não é significativamente grande, o que justifica a similaridade dos histogramas obtidos.

### **4.2.2.2 Análise da distribuição da variável premio\_exp por intensidade**

A fim de estudar a variação dos valores de premio\_exp numa análise de intensidade por localização, foram plotados no mapa cada ponto desta variável, obtendo-se os resultados abaixo:



**Figura 4: Distribuição pontual por intensidade dos valores de prêmio\_exp dos veículos Honda Civic no território nacional brasileiro**

**Fonte: Elaborada pela autora**

Os pontos verificados na figura 4 acima representam os valores de prêmio\_exp distribuídos por intensidade, de acordo com os valores a seguir, que constituem respectivamente, 1º, 2º e 3º quartil: R\$ 1.110,95; R\$ 1.412,78; R\$ 1.787,02.

Esta figura apresenta a localização dos valores de prêmio\_exp para o seguro do veículo Honda Civic, numa análise por intensidade de distribuição. Nas regiões onde estão localizados os pontos vermelhos, encontram-se os 25% maiores valores de prêmio para o seguro deste veículo, ou seja, estes pontos representam todos os valores

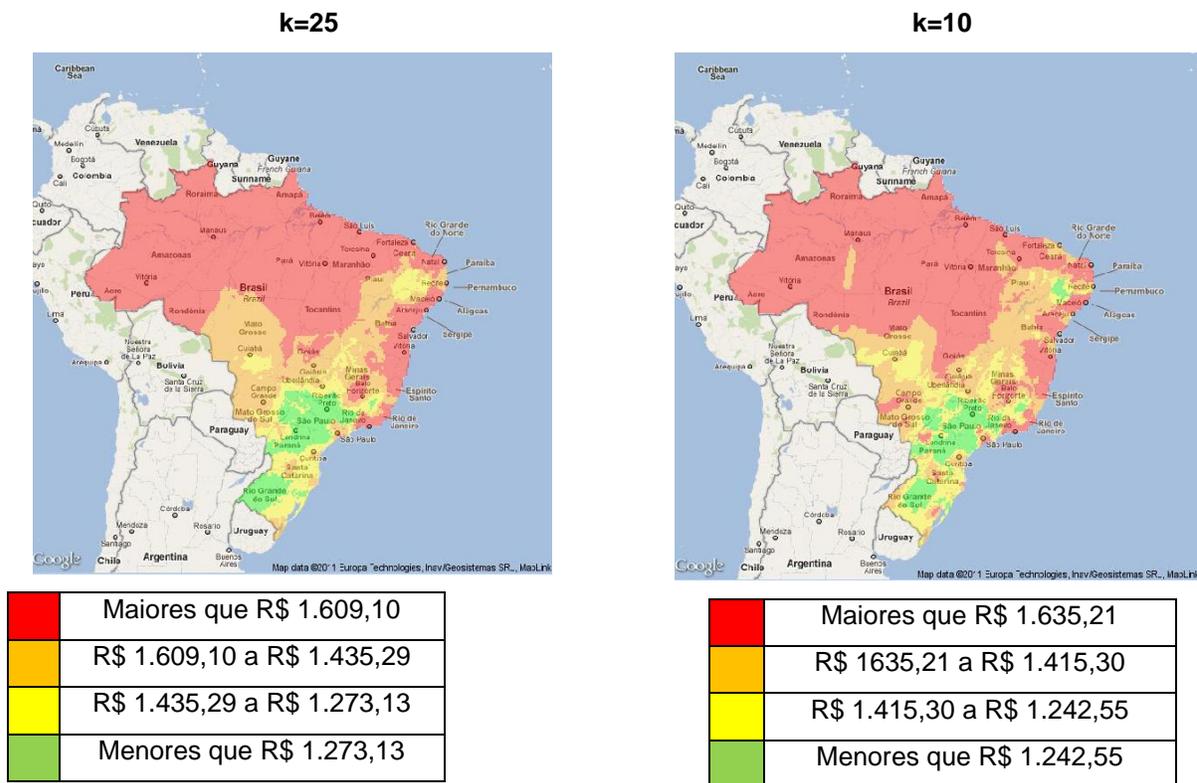
acima R\$ 1.787,02. Os pontos laranja referem-se à localização dos 75% menores valores da variável, representando todos os que estão abaixo do referido valor. Onde estão localizados os pontos amarelos, verificam-se os 50% menores valores, o que representa, portanto, os prêmios abaixo de R\$ 1.412,78. E finalmente, nos pontos verdes estão representados os 25% menores valores da variável, sendo então os prêmios abaixo de R\$ 1.110,95.

Desta forma, os menores valores de prêmio por exposição de um veículo Honda Civic podem ser encontrados com maior facilidade nas regiões Sudeste e Sul do Brasil, mais especificamente nos estados de São Paulo, Paraná e Rio Grande do Sul. Quanto aos maiores valores, pode-se localizá-los principalmente nos estados do Rio de Janeiro, Minas Gerais e Espírito Santo. Nas regiões Nordeste e Centro-oeste também se observam valores altos desta variável.

#### **4.2.2.3 Distribuição dos pontos por aproximação**

Para fins desta análise, primeiramente os resultados foram gerados através da comparação de cada ponto aos outros 25 pontos mais próximos ( $K=25$ ), em seguida, repetiu-se o procedimento comparando-se apenas aos 10 pontos mais próximos ( $K=10$ ), obtendo-se desta forma, dois resultados distintos.

Tendo em vista que esta análise está baseado em um conjunto de 1.326 pontos, tanto a comparação com  $k=25$  quanto a com  $k=10$  pontos foram consideradas sem o risco da obtenção de resultados pouco representativos, pois o total dos elementos é suficiente para se trabalhar com ambas as possibilidades.



**Figura 5: Comparação das suavizações com k=25 e k=10 da distribuição por intensidade dos valores de prêmio\_exp dos veículos Honda Civic**

**Fonte: Elaborada pela autora**

Como se pode observar através da figura 7, a suavização por aproximação dos valores feita na distribuição da variável prêmio\_exp gerou novas faixas de intensidade destes valores.

Na aproximação com os 25 pontos mais próximos, as faixas de intensidade resultantes baseiam-se nos valores R\$ 1.273,13; R\$ 1.435,29; R\$ 1.609,10, que constituem respectivamente, 1º, 2º e 3º quartil. Esta suavização mostra que há predominância dos maiores valores de prêmio\_exp para o seguro de um Honda Civic nas regiões Norte e Nordeste do país, verificando-se também esta ocorrência em parte da região Sudeste, mais especificamente no estado de Espírito Santo e nas capitais de Minas Gerais, Rio de Janeiro e São Paulo. Os menores valores estão localizados no interior dos estados de São Paulo, Rio de Janeiro, Paraná e Rio Grande do Sul, ou seja, em cidades interioranas das regiões Sul e Sudeste do país.

Valores médios-altos de k=25, coloração laranja, incidem principalmente na região Centro Oeste, interior do Nordeste, parte de Minas Gerais e nas capitais dos

estados do Sul. Quanto aos médios-baixos, coloração amarela, são essencialmente verificados na área metropolitana do Sul e demais cidades do Sudeste, com incidência considerável também em Cuiabá e cidades de Recife e Maceió.

Nesta aproximação, não se verifica em toda a região Sul e em quase totalidade do Sudeste, a presença de valores significativamente altos da variável em questão.

Com relação à aproximação com os 10 pontos mais próximos, as faixas de intensidade geradas baseam-se nos valores R\$ 1.242,55; R\$ 1.415,30; R\$ 1.635,21, que constituem respectivamente, 1º, 2º e 3º quartil.

Esta suavização apresenta-se com algumas diferenças em relação à anterior, que fazem com que o seu resultado fique mais próximo da configuração real dos dados representada na figura 4.

Observa-se na figura 7 que na aproximação com  $k=10$  os valores mais altos de prêmio\_exp incidem, principalmente, nas regiões Norte e Nordeste do país. Esta ocorrência é também verificada nas capitais do Sudeste e dos estados de Mato Grosso, Goiás e Rio Grande do Sul.

Assim como na aproximação com 25 pontos, os menores valores estão localizados no interior dos estados de São Paulo, Rio de Janeiro, Paraná e Rio Grande do Sul, ou seja, em cidades interioranas das regiões Sul e Sudeste do país. Nesta segunda aproximação, observa-se também esta incidência na região sertaneja, o que não foi verificado na primeira.

No interior da região Nordeste, na região Centro-Oeste e em grande parte do Sul e Sudeste, verifica-se a maior proporção dos valores médios de prêmio por exposição.

A ocorrência de valores altos na região Sul é extremamente reduzida, aumentando nas regiões Centro-Oeste, Sudeste, Nordeste e Norte, nesta ordem.

Com excessão do Norte do país, observa-se que os maiores valores estão situados principalmente nas capitais, dentre elas, somente Curitiba, Cuiabá, Aracaju e Fortaleza apresentam valores médios.

#### 4.2.2.4 Aplicação da Estatística de Varredura Espacial



**Figura 6: Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio\_exp para os veículos Honda Civic no território nacional brasileiro**

**Fonte: Elaborada pela autora**

Na aplicação do método de varredura espacial sobre as observações referentes aos veículos Honda Civic foram definidos apenas dois conjuntos, ou seja, na varredura dos dados foram identificados apenas dois conglomerados significativos, com p-valor abaixo de 0,05, sendo que o dos menores valores apresenta-se com p-valor igual 0,006 e o dos maiores com p-valor igual 0,004.

Pode-se observar que estes conglomerados estão situados nas regiões de maior concentração dos pontos, de acordo com a figura 4.

Desta forma, o conjunto dos maiores valores de prêmio\_exp abrange a região Nordeste e parte do Norte e Sudeste, aonde realmente se verifica a maior concentração dos valores mais altos. Já o conjunto dos valores menores desta variável compreende grande parte da região Sul, o estado de São Paulo e parte do Mato Grosso do Sul, ou

seja, as localidades aonde de fato se observa a maior incidência dos valores mais baixos de premio\_exp.

#### 4.2.3 Análise dos dados referentes aos modelos Fiat Uno 1.0

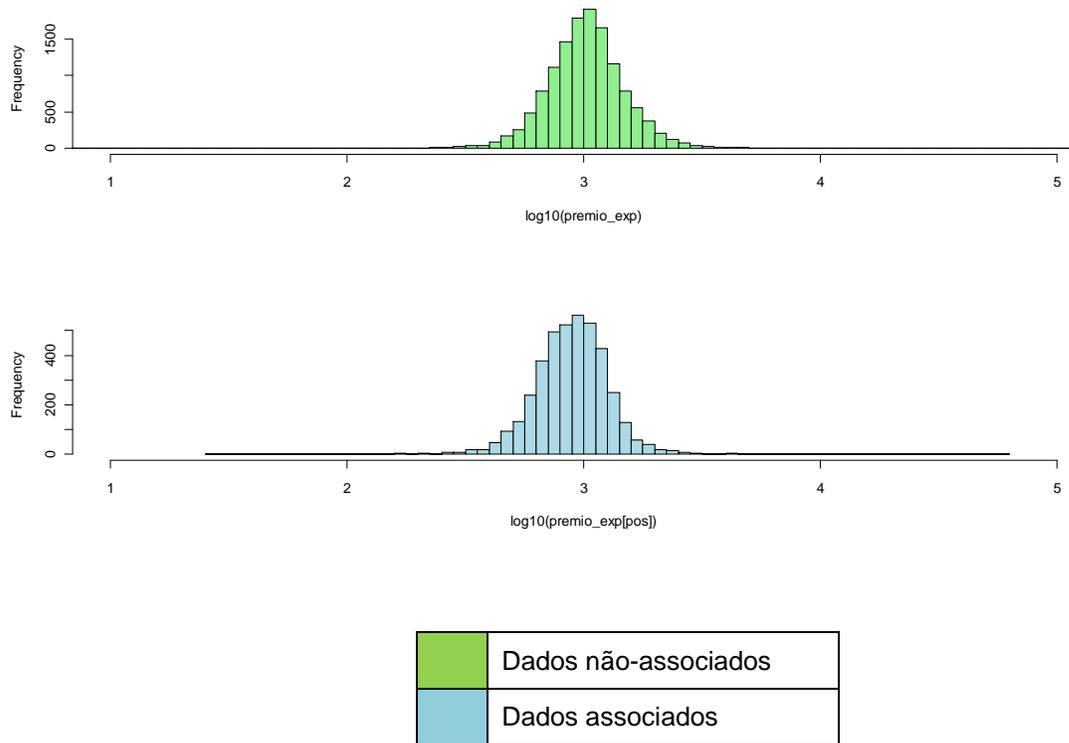


**Figura 7: Distribuição pontual das observações para os veículos Fiat Uno 1.0 no território nacional brasileiro**

**Fonte: Elaborada pela autora**

A figura 9 trata-se da representação da distribuição dos veículos Fiat Uno 1.0 no território brasileiro. A disposição dos pontos indica que há uma maior concentração destes veículos nas regiões Sul, Sudeste e Nordeste do país. Verifica-se também uma quantidade considerável de observações na porção Centro-Oeste. Já na região Norte, observa-se pouca incidência destes veículos.

#### 4.2.3.1 Análise descritiva da variável premio\_exp



**Figura 8: Histogramas da variável premio\_exp referente aos veículos Fiat Uno 1.0 antes e depois da associação das observações às coordenadas geográficas**  
**Fonte: Elaborada pela autora**

Os histogramas constantes na figura 10 representantes dos valores de premio\_exp referentes aos veículo Fiat Uno 1.0, respectivamente, antes e depois do vínculo dos dados às coordenadas geográficas, apresentam-se semelhantes, o que significa que, o histograma que representa todas as observações da base, aproxima-se do que representa apenas os dados para os quais foram localizadas as coordenadas. Isso indica que o número de dados associados, ou seja, a amostra obtida é satisfatória para o estudo, uma vez que sua representação aproxima-se do conjunto total dos dados.

## Análise descritiva dos dados antes da associação à base de cep's.

TABELA 3

Fiat Uno 1.0: dados antes da associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
1,50	763,90	986,30	1.168,00	1.264,00	453.700,00	91,10	0,8856	2,2e-16

Fonte: Elaborada pela autora

Através da figura 10 verifica-se que o histograma da variável premio\_exp referente aos dados não-associados apresenta-se relativamente simétrico.

Pelas estatísticas encontradas, observa-se que os valores desta variável giram em torno da média de R\$ 1.168,00 e que 50% deles estão acima de R\$ 986,30 e 50% estão abaixo deste valor.

Observa-se que o menor valor encontrado para esta variável é R\$ 1,50, sendo que 25% estão dispostos abaixo de R\$ 763,90 e 25% estão acima de R\$ 1.264,00. O valor máximo observado é igual a R\$ 453.700,00.

A verificação de um valor mínimo significativamente baixo (R\$ 1,5) e de um valor máximo notavelmente alto (R\$ 453.700,00) pode estar relacionada ao curto período de exposição analisado, que influencia na razão prêmio/exposição, ou a um possível erro de digitação de valores na base de dados em estudo.

O teste de normalidade aplicado indica que a variável premio\_exp referente aos dados não-associados não segue a distribuição normal, o que se pode verificar mediante ao baixo valor da estatística p-value, inferior a 0,05.

## Análise descritiva dos dados após a associação à base de cep's.

**TABELA 4**  
Fiat Uno 1.0: dados após a associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
1,53	811,90	1.022,00	1.147,00	1.269,00	120.400,00	14,00	0,9173	2,2e-16

Fonte: Elaborada pela autora

Observa-se na figura 10, que o histograma da variável premio\_exp referente aos dados associados também se apresenta relativamente simétrico. As estatísticas encontradas indicam que os valores desta variável giram em torno da média de R\$ 1.147,00, estando 50% deles dispostos acima de R\$ 1.022,00 e 50% abaixo. O valor mínimo encontrado para esta variável corresponde a R\$ 1,53.

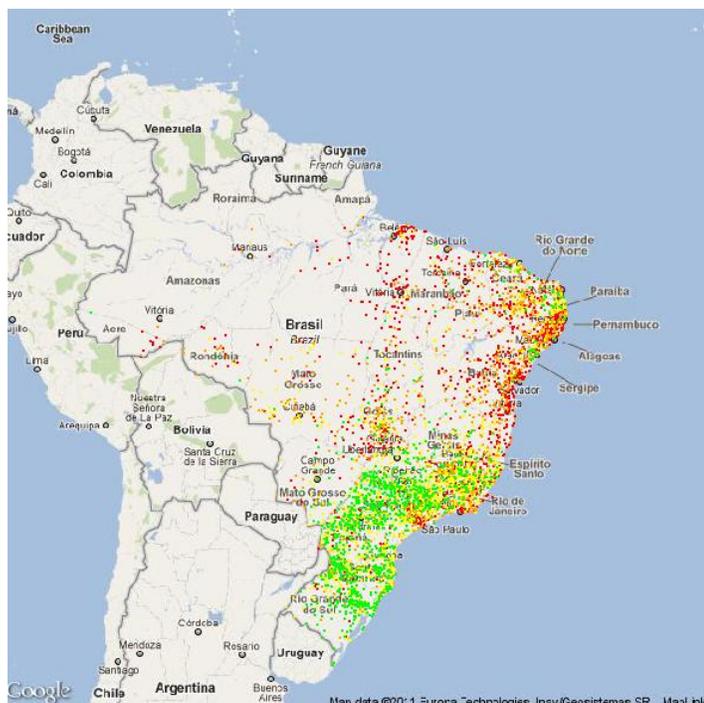
Verifica-se que 25% dos dados estão abaixo de R\$ 811,90 e 25% estão acima de R\$ 1.269,00, sendo o valor máximo encontrado igual a R\$ 120.400,00.

O resultado de um valor mínimo significativamente reduzido e de um valor máximo consideravelmente elevado justifica-se pelas mesmas razões apontadas na análise feita antes da associação à base de cep's.

O teste de normalidade aplicado indica que a variável premio por exposição referente aos dados associados não segue a distribuição normal, o que se pode verificar mediante ao baixo valor da estatística p-value, inferior a 0,05.

Pode ser observado que a diferença no resultado da maior parte das estatísticas encontradas em ambas as análises não é significativamente elevada, o que justifica a similaridade dos histogramas obtidos.

### **4.2.3.2 Análise da distribuição variável premio\_exp por intensidade**



**Figura 9: Distribuição pontual por intensidade dos valores de prêmio\_exp dos veículos Fiat Uno 1.0 no território nacional brasileiro**

**Fonte: Elaborada pela autora**

	Maiores que R\$ 1.111,15
	R\$ 1.111,15 a R\$ 897,51
	R\$ 897,51 a R\$ 718,29
	Menores que R\$ 718,29

Os pontos verificados na figura 11 representam os valores de prêmio\_exp distribuídos por intensidade, de acordo com os valores a seguir, que constituem respectivamente, 1º, 2º e 3º quartil: R\$ 718,29; R\$ 897,51; R\$ 1.111,15.

Esta figura apresenta a localização de cada faixa dos valores de prêmio\_exp para o seguro do veículo Fiat Uno 1.0, numa análise por intensidade da distribuição. Nas regiões onde estão localizados os pontos vermelhos, encontram-se os 25% maiores valores de prêmio para o seguro deste veículo, ou seja, estes pontos representam todos os valores acima R\$ 1.111,15. Os pontos laranja referem-se à localização dos 75% menores valores da variável, ou seja, todos os que estão abaixo deste valor. Onde estão localizados os pontos amarelos, verificam-se os 50% menores valores, representando, portanto, os prêmios abaixo de R\$ 897,51. E finalmente, nos

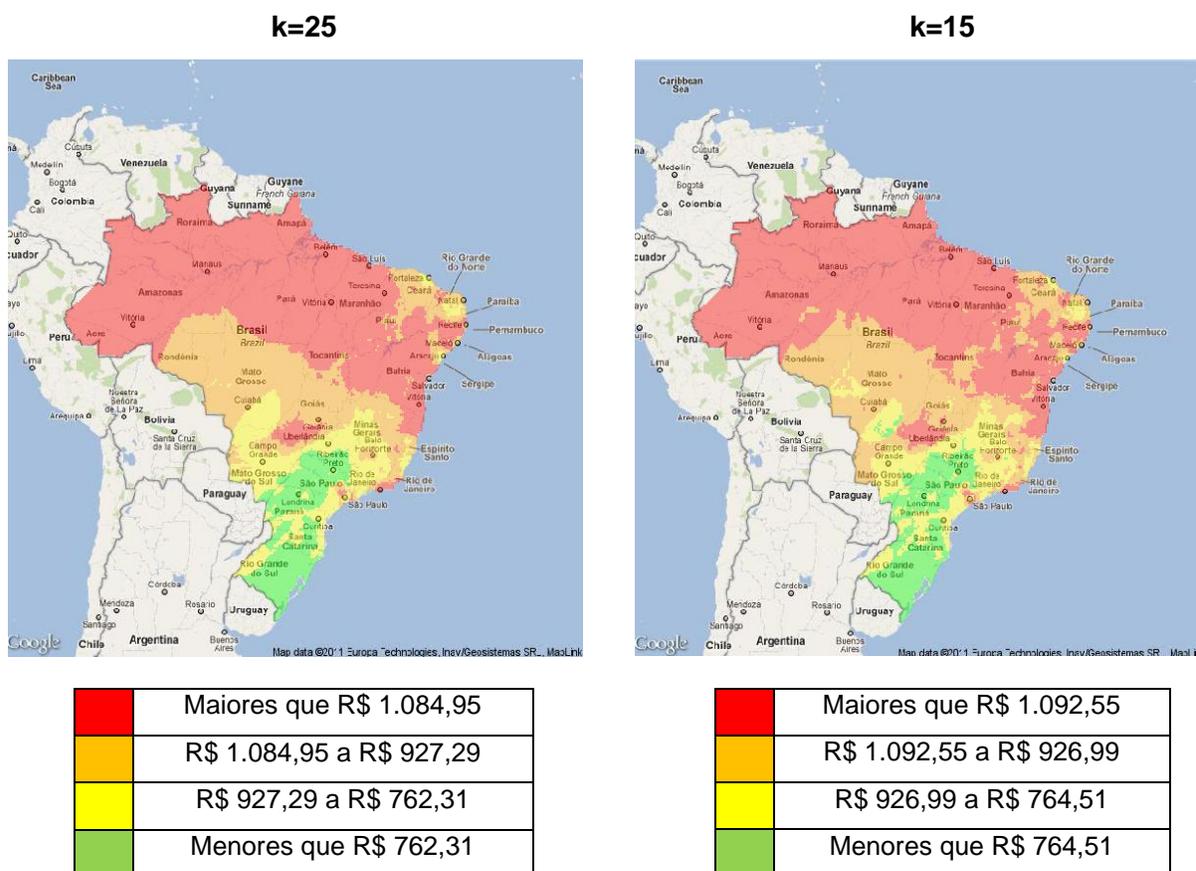
pontos verdes estão representados os 25% menores valores da variável, sendo então os prêmios abaixo de R\$ 718,29.

Desta forma, os menores valores de prêmio por exposição de um veículo Fiat Uno 1.0 podem ser encontrados com maior facilidade nas regiões Sudeste, principalmente nos estados de São Paulo, seguido de Rio de Janeiro e Minas Gerais e em toda a região Sul do Brasil. Quanto aos maiores valores, pode-se localizá-los principalmente no Nordeste, seguido do Centro-Oeste e Norte do país. Verifica-se a presença de valores elevados também no estado de Espírito Santo e nas grandes capitais do Sudeste.

#### **4.2.3.3 Distribuição dos pontos por aproximação**

Nesta análise, primeiramente os resultados foram gerados com o  $k=25$  e, em seguida com  $k=15$ .

Sabendo-se que este conjunto contém 4.031 pontos, tanto a aproximação com 25 quanto a com 15 pontos foram consideradas, sem o risco da obtenção de resultados pouco representativos, pois o total dos elementos é suficiente para se trabalhar com ambas as possibilidades.



**Figura 10: Comparação das suavizações com k=25 e k=15 da distribuição por intensidade dos valores de prêmio\_exp dos veículos Fiat Uno 1.0**

Fonte: Elaborada pela autora

Através da figura 14 é possível perceber que a suavização por aproximação dos valores feita na distribuição da variável prêmio\_exp gerou novas faixas de intensidade destes valores.

As faixas de intensidade resultantes na aproximação com os 25 pontos mais próximos baseam-se nos valores R\$ 762,31; R\$ 927,29; R\$ 1.084,95, que constituem respectivamente, 1º, 2º e 3º quartil. Esta suavização mostra que há predominância dos maiores valores de prêmio\_exp para o seguro de Fiat Uno 1.0 em toda a região Norte e na maior parte do Nordeste do país, verificando-se também esta ocorrência em uma porção do Centro-Oeste e nas capitais de Minas Gerais, Rio de Janeiro e São Paulo. Os menores valores estão localizados no interior do estado de São Paulo e na região Sul do país.

Valores médios-altos, coloração laranja, incidem principalmente na região Centro-Oeste, seguida dos estados do Ceará e parte de Minas Gerais. Quanto aos valores médios-baixos, coloração amarela, são essencialmente verificados em cidades do Sul, Centro-Oeste e Sudeste.

Em toda a região sul, não se observa a presença de valores altos da variável em estudo.

Relativamente à aproximação com os 15 pontos mais próximos, verifica-se que as faixas de intensidade geradas baseam-se nos valores R\$ 764,51; R\$ 926,99; R\$ 1.092,55, que constituem respectivamente, 1º, 2º e 3º quartil.

A diferença entre a aproximação com os 15 pontos mais próximos e a anterior é pequena. Verifica-se apenas que a suavização com  $k=15$  gerou uma distribuição um pouco menos homogênea do que a anterior, mas as características regionais marcantes permanecem as mesmas.

#### 4.2.3.4 Aplicação da Estatística de Varredura Espacial



**Figura 11: Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio\_exp para os veículos Fiat Uno 1.0 no território nacional brasileiro**

**Fonte: Elaborada pela autora**

Para os veículos Fiat Uno 1.0, assim como para os Honda Civic, na aplicação do método de varredura espacial foram definidos apenas dois conjuntos, ou seja, na varredura dos dados foram definidos apenas dois conglomerados significativos, com p-valor abaixo de 0,05, sendo que tanto o dos menores valores quanto o dos maiores apresentam-se com p-valor igual 0,001.

Observa-se que os conglomerados formados estão situados nas regiões de maior concentração dos pontos, de acordo com a figura 11.

Sendo assim, o conjunto dos maiores valores de prêmio\_exp abrange a região Nordeste e parte do Norte e Sudeste, aonde realmente se verifica a maior concentração dos valores mais altos. Da mesma forma, o grupo dos valores menores desta variável está situado nas localidades aonde de fato se observa a maior incidência dos valores

mais baixos de `premio_exp`, compreendendo então, toda a região Sul, o estado de São Paulo, parte do Rio de Janeiro e do Mato Grosso do Sul.

#### 4.2.4 Análise dos dados referentes aos modelos Fiat Siena 1.0

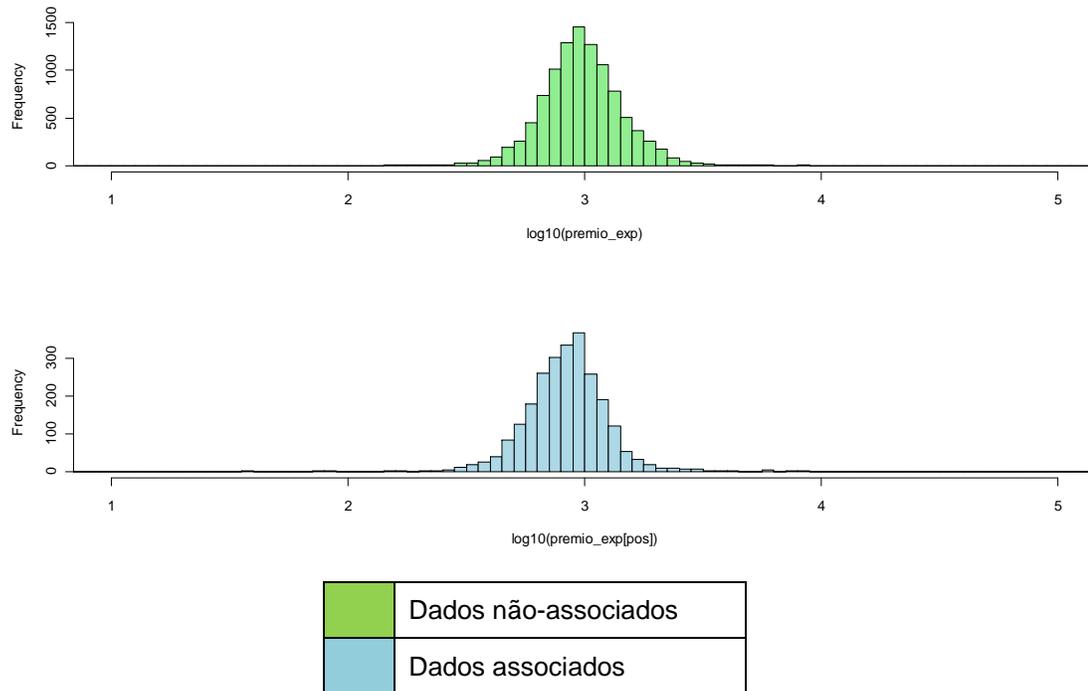


**Figura 12: Distribuição pontual das observações para os veículos Fiat Siena 1.0 no território nacional brasileiro**

**Fonte: Elaborada pela autora**

A figura 16 representa a distribuição dos veículos pertencentes ao modelo Fiat Siena 1.0. A disposição dos pontos indica que há uma maior concentração destes veículos nas regiões Sul, Sudeste e Nordeste do país. Mas verifica-se também uma quantidade razoável de pontos em parte da porção Centro-Oeste, enquanto que na região Norte encontram-se somente poucas observações isoladas.

#### 4.2.4.1 Análise descritiva da variável premio\_exp



**Figura 13: Histogramas da variável premio\_exp referente aos veículos Fiat Siena 1.0 antes e depois da associação das observações às coordenadas geográficas**

Fonte: Elaborada pela autora

Os histogramas constantes na figura 17, representantes dos valores de premio\_exp referentes aos veículos Fiat Siena 1.0, respectivamente, antes e depois do vínculo dos dados às coordenadas geográficas, apresentam-se relativamente semelhantes. A diferença verificada entre ambos não é considerável, o que significa que, o histograma que representa todos os dados da base relativos ao Fiat Siena 1.0, aproxima-se do que representa apenas os dados deste modelo para os quais foram localizadas as coordenadas. Isso indica que o número de dados associados, ou seja, a amostra obtida é satisfatória para o estudo, uma vez que sua representação aproxima-se do conjunto total dos dados.

### Análise descritiva dos dados antes da associação à base de cep's.

**TABELA 5**  
Fiat Siena 1.0: dados antes da associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
2,10	750,90	959,50	1.225,00	1.245,00	496.400,00	32,00	0,8697	2,2e-16

Fonte: Elaborada pela autora

A figura 17 mostra que o histograma da variável premio\_exp referente aos dados não-associados apresenta-se simétrico.

Através das estatísticas encontradas, observa-se que os valores desta variável giram em torno da média de R\$ 1.225,00 e que 50% deles estão acima de R\$ 959,50 e 50% abaixo deste valor.

Verifica-se que o menor valor encontrado para esta variável é R\$ 2,10, sendo que 25% estão dispostos abaixo de R\$ 750,90 e 25% estão acima de R\$ 1.245,00. O valor máximo observado é igual a R\$ 496.400,00.

A observação de um valor mínimo significativamente baixo (R\$ 2,10) e de um valor máximo notavelmente alto (R\$ 496.400,00) pode estar relacionada ao curto período de exposição analisado, que influencia na razão prêmio/exposição, ou a um possível erro de digitação de valores na base de dados em estudo.

O teste de normalidade aplicado indica que a variável premio\_exp referente aos dados não-associados não segue a distribuição normal, o que se pode verificar mediante ao baixo valor da estatística p-value, inferior a 0,05.

### Análise descritiva dos dados após a associação à base de cep's.

**TABELA 6**  
Fiat Siena 1.0: dados após a associação à base de cep's

Estatísticas Descritivas							Teste de normalidade	
Mín.	1° quartil	Médiana	Média	3° quartil	Máx.	Na's	W	p-valor
3,06	768,30	963,20	1.189,00	1.228,00	254.400,00	20,00	0,8568	2,2e-16

Fonte: Elaborada pela autora

Através da figura 17, observa-se que o histograma da variável premio\_exp referente aos dados associados também se apresenta relativamente simétrico. As estatísticas encontradas indicam que os valores desta variável giram em torno da média de R\$ 1.189,00, estando 50% deles dispostos acima de R\$ 963,20 e 50% abaixo. O valor mínimo encontrado para esta variável é R\$ 3,06.

Verifica-se que 25% dos dados estão abaixo de R\$ 768,30 e 25% estão acima de R\$ 1.228,00, sendo o valor máximo encontrado igual a R\$ 254.400,00.

O resultado de um valor mínimo significativamente reduzido e de um valor máximo consideravelmente elevado justifica-se pelas mesmas razões apontadas na análise feita antes da associação à base de cep's.

O teste de normalidade aplicado indica que a variável premio por exposição referente aos dados associados não segue a distribuição normal, o que se pode verificar mediante ao baixo valor da estatística p-value, inferior a 0,05.

Pode ser observado que a diferença no resultado da maior parte das estatísticas encontradas em ambas as análises não é significativamente elevada, o que justifica a similaridade dos histogramas obtidos.

#### **4.2.4.2 Análise da distribuição da variável premio\_exp por intensidade**



**Figura 14: Distribuição pontual por intensidade dos valores de prêmio\_exp dos veículos Fiat Siena 1.0 no território nacional brasileiro**

Fonte: Elaborada pela autora

	Maiores que R\$ 1.037,62
	R\$ 1.037,62 a R\$ 846,28
	R\$ 846,28 a R\$ 664,43
	Menores que R\$ 664,43

Os pontos verificados na figura 18 representam os valores de prêmio\_exp distribuídos por intensidade, de acordo com os valores a seguir, que constituem respectivamente, 1º, 2º e 3º quartil: R\$ 664,43; R\$ 846,28; R\$ 1.037,62.

Esta figura apresenta a localização de cada faixa dos valores de prêmio\_exp para o seguro do veículo Fiat Siena 1.0, numa análise por intensidade da distribuição. Nas regiões onde estão localizados os pontos vermelhos, encontram-se os 25% maiores valores de prêmio para o seguro deste veículo, ou seja, estes pontos representam todos os valores acima R\$ 1.037,62. Os pontos laranja referem-se à localização dos 75% menores valores da variável, ou seja, todos os que estão abaixo deste valor. Onde estão localizados os pontos amarelos, verificam-se os 50% menores valores, representando, portanto, os prêmios abaixo de R\$ 846,28. E finalmente, nos

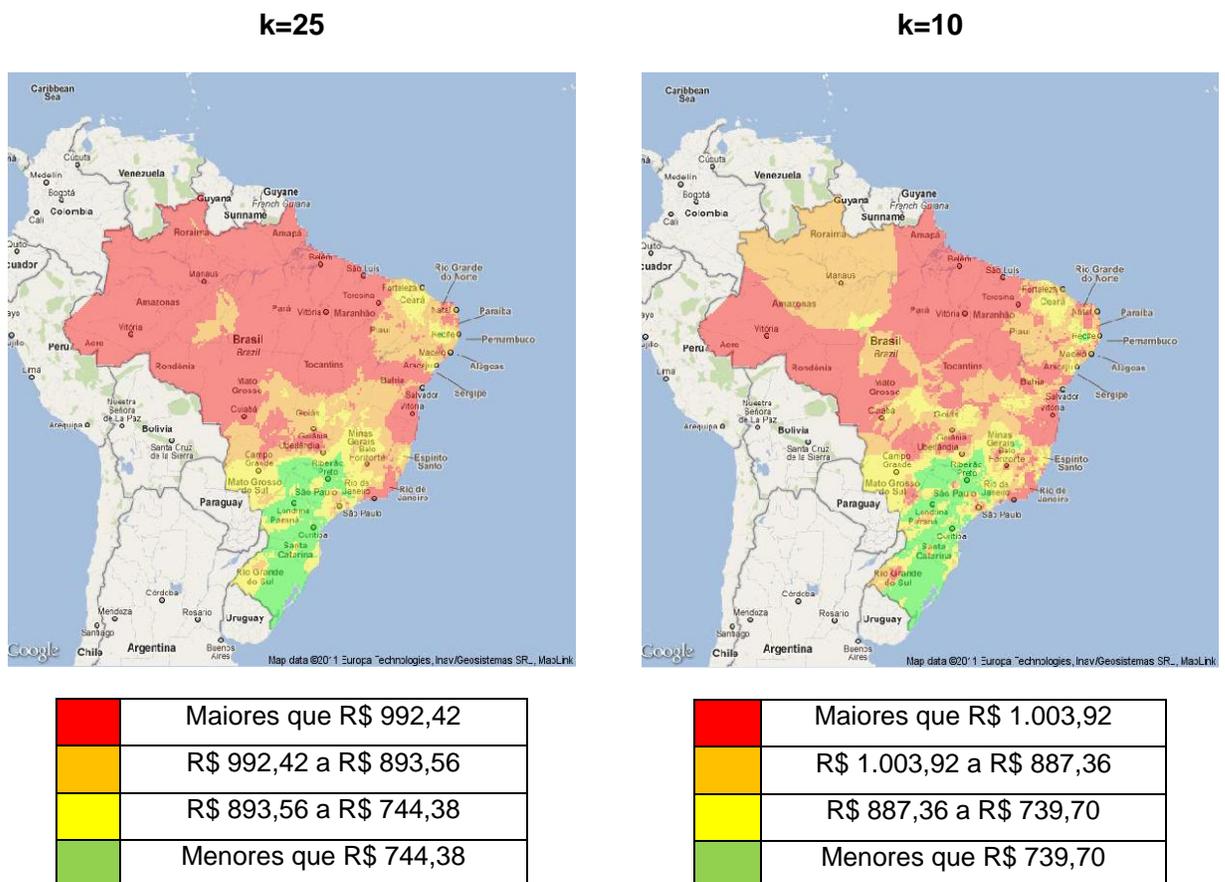
pontos verdes estão representados os 25% menores valores da variável, sendo então os prêmios abaixo de R\$ 664,43.

Desta forma, os menores valores de prêmio por exposição de um veículo Fiat Siena 1.0 podem ser encontrados com maior facilidade nas regiões Sul e Sudeste, principalmente no estado de São Paulo. Verifica-se também uma quantidade razoável destes pontos distribuídos no Nordeste do Brasil. Quanto aos maiores valores, pode-se localizá-los com maior evidência principalmente nas capitais dos estados de São Paulo e Rio de Janeiro. Há também uma grande concentração de valores altos em Minas Gerais, Espírito Santo e nas regiões Nordeste e Centro-Oeste.

#### **4.2.4.3 Distribuição dos pontos por aproximação**

Para fins desta análise, primeiramente os resultados foram gerados com o  $k=25$  e em seguida com o  $k=10$ .

Sabendo-se que este conjunto contém 2.529 pontos, tanto a aproximação com 25 quanto a com 10 pontos foram consideradas, sem o risco da obtenção de resultados pouco representativos, pois o total dos elementos é suficiente para se trabalhar com ambas as possibilidades.



**Figura 15: Comparação das suavizações com  $k=25$  e  $k=10$  da distribuição por intensidade dos valores de prêmio\_exp dos veículos Fiat Siena 1.0**

**Fonte: Elaborada pela autora.**

Por meio da figura 21, é possível observar que a suavização por aproximação dos valores feita na distribuição da variável prêmio\_exp gerou novas faixas de intensidade destes valores.

Na aproximação com os 25 pontos mais próximos, as faixas de intensidade resultantes baseam-se nos valores R\$ 744,38; R\$ 893,56; R\$ 992,42, que constituem respectivamente, 1º, 2º e 3º quartil. Esta suavização mostra que há predominância dos maiores valores de prêmio\_exp para o seguro de um Fiat Siena 1.0 em toda a região Norte. Observa-se também estes valores com significativa evidência em parte do Nordeste, Centro-Oeste e Rio de Janeiro.

Os menores valores estão localizados na região Sul do país e no interior do estado de São Paulo.

Valores médios-altos, coloração laranja, incidem principalmente nas regiões Centro-Oeste, Sudeste e Nordeste. Quanto aos valores médios-baixos, coloração amarela, são verificados com maior facilidade em parte do Sul, Centro-Oeste, Sudeste e Nordeste do Brasil.

Não se verifica em toda a região sul, a presença de valores altos da variável premio\_exp.

Sobre a aproximação com os 10 pontos mais próximos, as faixas de intensidade geradas baseam-se nos valores R\$ 739,70; R\$ 887,36; R\$ 1.003,92, que constituem respectivamente, 1º, 2º e 3º quartil.

A diferença entre a aproximação com os 10 pontos mais próximos e a anterior é visível com maior evidência apenas nas regiões Norte e Centro-Oeste do país. Isso porque, por utilizar-se de uma menor quantidade de pontos na aproximação, verifica-se que a suavização com  $k=10$  gerou uma distribuição menos homogênea do que a anterior, mas as características regionais marcantes permanecem as mesmas.

#### **4.2.4.4 Aplicação da Estatística de Varredura Espacial**



**Figura 16: Localização espacial dos conglomerados de menor (azul) e maior (vermelho) valor de prêmio\_exp para os veículos Fiat Siena 1.0 no território nacional brasileiro**

**Fonte: Elaborada pela autora**

A aplicação do método de varredura espacial para os modelos Fiat Siena 1.0, assim como para os Honda Civic e Fiat Uno 1.0, gerou apenas dois conjuntos, ou seja, na varredura dos dados foram definidos apenas dois conglomerados significativos, com p-valor abaixo de 0,05, sendo que o dos menores valores apresenta-se com p-valor igual 0,001 e o dos maiores com p-valor igual 0,002.

Verifica-se que os conglomerados formados se encontram dispostos nas regiões de maior concentração dos pontos, de acordo com a figura 18.

Nestas condições, o conjunto dos maiores valores de prêmio\_exp abrange a região Nordeste e parte do Norte e Sudeste, aonde realmente se verifica a maior concentração dos valores mais altos. Já o conjunto dos valores menores desta variável compreende a região Sul, a maior parte do estado de São Paulo e uma pequena porção do Mato Grosso do Sul e Rio de Janeiro, ou seja, as localidades aonde de fato se observa a maior ocorrência dos menores valores de prêmio\_exp.

## 5 CONCLUSÃO

Diante do objetivo de identificação de conglomerados espaciais para os valores de prêmio de seguros de veículos no território nacional brasileiro, verifica-se que a metodologia aplicada foi eficaz no alcance desta meta.

As técnicas utilizadas possibilitaram visualizar a distribuição do custo do seguro de automóvel no território brasileiro, apontando aonde se pode encontrar os maiores e menores valores cobrados pela prestação deste serviço no país.

A análise dos resultados encontrados para todos os veículos que serviram de base neste estudo levaram à conclusões semelhantes, permitindo especulações acerca dos motivos que influenciam a verificação da distribuição definida.

Em todas as análises foi possível observar que a região Sul é a que apresenta os menores valores de prêmio e que os maiores valores são identificados no Nordeste, nas capitais de cidades do Sudeste e no Centro-Oeste do país. Observando-se que no Sudeste e Centro-Oeste há uma maior diversificação dos valores do serviço em questão. Quanto à porção Norte, as conclusões obtidas basearam-se em resultados aproximados tendo em vista as poucas observações da base de dados situadas nesta região. Tal procedimento levou à verificação de forte incidência de valores altos nesta porção do território brasileiro.

Baseando-se nos critérios de determinação do prêmio em seguros de veículos e sabendo-se que o índice de sinistralidade diminui em função de diversos fatores, tais como, maior organização urbana, população reduzida, melhor distribuição de renda e menor índice de criminalidade, é possível afirmar que a presença de valores elevados de prêmio nas grandes capitais pode estar relacionada à alta sinistralidade comumente presente nestas áreas. Quanto à verificação de valores mais baixos, principalmente no Sul e regiões interioranas, pode-se dizer que tal situação é verificada em função da maior observância dos critérios redutores da sinistralidade nestas regiões.

## REFERÊNCIAS

- BRASIL, Ministério da Fazenda. Superintendência de Seguros Privados. **Base de dados do Autoseg**, atualizada até 1º semestre de 2010 (gerada em 04/04/2011 16:00:00). Disponível em: < <http://www.susep.gov.br/downloads.asp> >. Acesso em: 15 jun. 2011.
- FERREIRA, Paulo Pereira. **Modelos de precificação e ruína para seguros de curto prazo**. Rio de Janeiro: FUNENSEG, 2002. 210p.
- FUNDAÇÃO ESCOLA NACIONAL DE SEGUROS. Diretoria de Ensino e Pesquisa. **Teoria geral do seguro I**. Rio de Janeiro: FUNENSEG, 2005. 100p.
- IRB – BRASIL RE. Antônio Lober Ferreira de Souza. **Dicionário de seguros**: vocabulário conceituado de seguros. Rio de Janeiro: FUNENSEG, 2000. 160p.
- KULLDORFF, Martin; HUANG, Lan; KONTY, Kevin. A scan statistic for continuous data based on the normal probability model. **International Journal of Health Geographics**, Estados Unidos, v. 8, n. 58, out. 2009. Disponível em: < <http://www.ij-healthgeographics.com/content/8/1/58> >. Acesso em: 15 jul. 2011.
- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003. 311p.
- MBI. **Ceps e coordenadas geográficas**. Disponível em: < [www.mbi.com.br](http://www.mbi.com.br) >. Acesso em: 30 jul. 2011.
- MINGOTI, Sueli Aparecida. **Análise de dados através de métodos de estatística multivariada**: uma abordagem aplicada. Belo Horizonte: Ed. UFMG, 2005. 295p. (Didática ;8 ) ISBN 857041451X
- MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos, c2009. xvi, 493 p.
- MOORE; WEATHERFORD. **Simulação de Monte Carlo**. Risk Management, 2005. Disponível em: < <http://risk.nuvvo.com/lesson/5864-simulacao-de-monte-carlo> >. Acesso em: 19 ago. 2011.
- PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS. Coordenação do Curso de Ciências Atuariais. **Teoria do risco I**. Belo Horizonte, 2007. 25p.
- PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS. Pró-Reitoria de Graduação. Sistema de Bibliotecas. **Padrão PUC Minas de normalização**: normas da ABNT para apresentação de trabalhos científicos, teses, dissertações e monografias. Belo Horizonte, 2010. 52p. Disponível em <

[http://www.pucminas.br/documentos/normalizacao\\_projetos.pdf](http://www.pucminas.br/documentos/normalizacao_projetos.pdf)>. Acesso em: 15 ago. 2011.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. 4. ed. São Paulo: Atlas, 2003. 96p.

## ANEXO

### COMANDOS UTILIZADOS NO SOFTWARE ESTATÍSTICO R PARA LEITURA DA BASE DE DADOS DA SUSEP, ANÁLISE DESCRITIVA DAS OBSERVAÇÕES E APLICAÇÃO DAS TÉCNICAS DE SUAVIZAÇÃO E VARREDURA ESPACIAL

#### LEITURA DA BASE DE DADOS:

```
## Data: 20/05/2011, atualizado em 15/06/2011
## Programa para leitura da Base de Dados da SUSEP
```

```
rm(list=ls(all=TRUE))
require(RODBC)
```

```
#setwd("C:\\disciplinas\\Tarifacao\\R\\Geo_SUSEP")
```

```
setwd("C:\\disciplinas\\Tarifacao\\dados\\")
```

```
# Connecta ao Banco de Dados do tipo ACCESS
con <- odbcConnectAccess("baseAuto.mdb")
```

```
#####
```

#### SELEÇÃO DE VEÍCULOS POR CÓDIGO:

```
consulta <- sqlQuery(con, paste("select arq_casco3_comp.CEP,",
"arq_casco3_comp.EXPOSICAO, arq_casco3_comp.PREMIO,",
"arq_casco3_comp.FREQ_SIN1, arq_casco3_comp.INDENIZ1",
"from arq_casco3_comp",
"inner join auto2_vei",
"on arq_casco3_comp.COD_MODELO = auto2_vei.CODIGO",
"where arq_casco3_comp.ANO_MODELO = '2010' and auto2_vei.COD_GRUPO = 112"))
```

```
dim(consulta)
```

```
# Encerra a conexao
```

```
odbcClose(con)
```

```
#####
```

#### ANÁLISE DESCRITIVA DE TODOS OS DADOS REFERENTES AO VEÍCULO SOLICITADO CONSTANTE NA BASE:

```
##dt<-(consulta[,2:5])
```

#### #Geração da Variável premio\_exp#

```
##premio_exp <- dt$PREMIO/dt$EXPO
```

```
##summary(premio_exp)
```

```
##hist(log10(premio_exp), breaks=30, prob=TRUE, main="PREMIO_EXPOSIÇÃO DENSITY", xlab =
"PREMIO_EXPOSIÇÃO", ylab="DENSITY")
```

#### Teste de Normalidade:

```
##shapiro.test(sample(log10(premio_exp),size=5000))
```

#### ## GERANDO DADOS AGREGADOS (I)

```
dt <- aggregate(consulta[,2:5],list(CEP=consulta$CEP),sum)
```

```
## fix(dt)
```

```
#####
```

```
ASSOCIA OS CEPs ÀS COORDENADAS GEOGRÁFICAS
```

```
#####
```

```

geo_data <- read.csv2("CEPs_GEO(b).dat")
geo_data$lat <- as.numeric(as.character(geo_data$lat))
geo_data$lon <- as.numeric(as.character(geo_data$lon))
dt$lat <- NA
dt$lon <- NA
## merge
## aux <- 1:length(dt$CEP)      ## Determina se distâncias podem ser exploradas
for( cont in 1:length(dt$CEP) ){
pos <- geo_data$newCEP == dt$CEP[cont]
## aux[cont] <- min( abs(geo_data$newCEP - dt$CEP[cont]) )
if(sum(pos)>=1){
pos <- which.max(as.integer(pos))
dt$lat[cont] <- geo_data$lat[pos]
dt$lon[cont] <- geo_data$lon[pos]
##cat("\n",dt$CEP[cont])
}
}
## Número de pontos NÃO faltantes
sum(!is.na(dt$lat))          ## [1] 3268
#####
## COMPARA OS HISTOGRAMAS DOS PREMIOSEXPOSICAO ANTES E DEPOIS DO VINCULO GEO

```

#### Geração da Variável premio\_exp:

```

premio_exp <- dt$PREMIO/dt$EXPO
pos <- !is.na(dt$lat)
summary(premio_exp)
hist(log10(premio_exp), breaks=30, prob=TRUE, main="PREMIO_EXPOSIÇÃO DENSITY", xlab =
"PREMIO_EXPOSIÇÃO", ylab="DENSITY")
windows()
par(mfrow=c(2,1))
hist(log10(premio_exp),breaks=100,main="",col='light green',xlim=c(1,5))
hist(log10(premio_exp[pos]),breaks=100,main="",col='light blue',xlim=c(1,5))

```

#### Teste de Normalidade:

```

##shapiro.test(sample(log10(premio_exp),size=5000))
#####
## RETIRA OS DADOS FALTANTES
#####
pos <- !is.na(dt$lat)
dt <- dt[pos,]
#####
## USE GOOGLE MAPS TO EXPLORE THE DATA
#####
## Plota os polígonos com os códigos de cores pré-definidos
#####
require(RgoogleMaps)
require(PBSmapping)
zoom <- 4
maptype <- 'roadmap' ##'terrain', 'roadmap', 'satellite', 'hybrid'
n_pix <- 640
center <- c(lat=-11.126663, lon=-51.533207)
destfile <- 'TemporaryMap.jpg'

BrasilMap <- GetMap(center = center[c('lat','lon')],
size = c(n_pix, n_pix), zoom = zoom, format = 'jpg',
maptype = maptype, destfile = destfile)

```

```

windows()
PlotOnStaticMap(BrasilMap)
#####
## PLOTA OS PONTOS REFERENTES AO VEICULO SELECIONADO##
PlotOnStaticMap(BrasilMap,dt$lat,dt$lon,
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='red')

## PLOTA TODOS OS PONTOS NO MAPA
PlotOnStaticMap(BrasilMap,geo_data$lat,geo_data$lon,
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='blue')
#####
write.table(dt,"C:\\disciplinas\\Tarifacao\\dados\\SUSEP_GEO.dat",quote=F,col.names=T,row.names=F,se
p='\t')
## dt <- read.table("C:\\disciplinas\\Tarifacao\\dados\\SUSEP_GEO.dat",header=T,sep=")
#####
## Carrega os limites do Território Brasileiro, obtido em:
## http://www.gadm.org/

BRAnew <- read.table("C:\\disciplinas\\Tarifacao\\dados\\BRAnew.dat",header=T,sep="")
#####
## DESENHA OS PONTOS SEGUNDO "INTENSIDADE"
pos <- dt$EXPO != 0
dt <- dt[pos,]
premio_exp <- dt$PREMIO/dt$EXPOSICAO
log10_prem_exp <- log10(premio_exp)
## log10_prem_exp <- log10(premio_exp+1)
aux <- quantile(log10_prem_exp,prob=c(0.25,0.5,0.75))
PlotOnStaticMap(BrasilMap)

pos <- (log10_prem_exp >= aux[3])
PlotOnStaticMap(BrasilMap,dt$lat[pos],dt$lon[pos],
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='red')

pos <- (log10_prem_exp >= aux[2])&(log10_prem_exp < aux[3])
PlotOnStaticMap(BrasilMap,dt$lat[pos],dt$lon[pos],
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='orange')

pos <- (log10_prem_exp >= aux[1])&(log10_prem_exp < aux[2])
PlotOnStaticMap(BrasilMap,dt$lat[pos],dt$lon[pos],
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='yellow')

pos <- (log10_prem_exp < aux[1])
PlotOnStaticMap(BrasilMap,dt$lat[pos],dt$lon[pos],
add=TRUE,destfile=destfile,cex=0.3,pch=20,col='green')

cat("Premio/Exposicao (pontos): ",10^(aux), "\n")

#####
## DESENHA AS INTENSIDADES POR SUAVIZAÇÃO##
require('spatialkernel') ## Para testar se um ponto esta no poligono
windows()
PlotOnStaticMap(BrasilMap)

## The 'QUICK and DIRTY' method (assuming the Earth is a perfect sphere)
const <- 180/pi
R <- 6378.7 ##(km)

```

```

x <- dt$lon*60*1852*cos(dt$lat/const)
y <- dt$lat*60*1852

## ----- ##
## PARAMETROS DE GRID ##
## ----- ##
## names(BRA)
gr <- 200 ## 200
k <- 25 ## k-nearest-neighbors : 25, 15, 100
xrange <- c(min(BRAnew$lon), max(BRAnew$lon))
yrange <- c(min(BRAnew$lat), max(BRAnew$lat))
mrate <- matrix(NA,gr,gr)
xseq <- seq(xrange[1],xrange[2],by=(xrange[2]-xrange[1])/(gr-1))
yseq <- seq(yrange[1],yrange[2],by=(yrange[2]-yrange[1])/(gr-1))

## -----
## PRIMEIRO AVALIA A INTENSIDADE NOS PONTOS OBSERVADOS PARA DEFINIR A ESCALA DO
MAPA ## -----
intensidade <- c()
for(cont in 1:length(dt$lon)){
x0 <- dt$lon[cont]*60*1852*cos(dt$lat[cont]/const)
y0 <- dt$lat[cont]*60*1852

dis <- (x0-x)^2 + (y0-y)^2 ## MAIS SIMPLES
aux <- order(dis)[1:(k+1)] ## ORDEM

intensidade <- c( intensidade, sum(dt$PREMIO[aux])/sum(dt$EXPOSICAO[aux]) )
}

## windows(); plot((dt$PREMIO/dt$EXPOSICAO), intensidade, xlab='taxas brutas', ylab='knn')

## Define os quartis para as definicoes de cores: vermelho, laranja, amarelo, verde
range <- quantile(intensidade,probs=c(0.25,0.5,0.75))
cat("Premio/Exposicao (suavizado): ",range, "\n")

## -----
## REALIZA A VARREDURA NO GRID
## -----

cx <- 1
for(lon0 in xseq){
cy <- 1
for(lat0 in yseq){
## VERIFICA SE O PONTO ESTÁ DENTRO DO POLIGONO DO BRASIL
if( pinpoly(cbind(BRAnew$lon,BRAnew$lat), cbind(lon0,lat0)) > 0 ){
x0 <- lon0*60*1852*cos(lat0/const)
y0 <- lat0*60*1852

dis <- (x0-x)^2 + (y0-y)^2 ## MAIS SIMPLES
aux <- order(dis)[1:(k+1)] ## ORDEM

mrate[cx,cy] <- sum(dt$PREMIO[aux])/sum(dt$EXPOSICAO[aux])
## ## mrate[cx,cy] <- mean(premio_exp[aux])
}
cy <- cy + 1
}
}

```

```

cx <- cx + 1
}

## -----

xgrid <- (xrange[2]-xrange[1])/(gr-1)
ygrid <- (yrange[2]-yrange[1])/(gr-1)
cx <- 1
for(lon0 in xseq){
cy <- 1
for(lat0 in yseq){

## VERIFICA SE O PONTO ESTÁ DENTRO DO POLIGONO DO BRASIL
if( !is.na(mrate[cx,cy]) ){

poly <- data.frame(PID=rep(1,5),POS=c(1:5),
X=c(lon0,lon0+xgrid,lon0+xgrid,lon0,lon0) - (xgrid/2),
Y=c(lat0,lat0,lat0-ygrid,lat0-ygrid,lat0) + (ygrid/2))

poly <- as.PolySet(poly,projection='LL')

if( mrate[cx,cy] < range[1] ){ col = "#00FF0066" }
if((mrate[cx,cy] >= range[1])&(mrate[cx,cy] < range[2])){ col = "#FFFF0066" }
if((mrate[cx,cy] >= range[2])&(mrate[cx,cy] < range[3])){ col = "#FFA50066" }
if( mrate[cx,cy] >= range[3]){ col = "#FF000066" }

PlotPolysOnStaticMap(BrasilMap, poly, lwd=.25, col=col, add=T, border=NA)

}
cy <- cy + 1
}
cx <- cx + 1
}

## windows(); image(-mrate)
## hist(mrate,col='light yellow')

#####
## GERA OS ARQUIVOS DE ENTRADA PARA O APLICATIVO SATSCAN
#####
INDEX <- 1:length(log10_prem_exp)
ONES <- 1+0*INDEX
write.table(cbind(INDEX,ONES,log10_prem_exp), "premio_exposicao.cas", quote=F, sep="\t",
col.names=F,row.names=F)

write.table(cbind(INDEX,dt$lat,dt$lon), "cordenadas.geo", quote=F, sep="\t",
col.names=F,row.names=F)

#####
## EXECUTA O SATSCAN PARA DEFINIR OS CLUSTERS DE ALTA INTENSIDADE
system('SaTScanBatch.exe high_satscan.prm',show.output.on.console = TRUE)

## EXECUTA O SATSCAN PARA DEFINIR OS CLUSTERS DE BAIXA INTENSIDADE
system('SaTScanBatch.exe low_satscan.prm',show.output.on.console = TRUE)
#####

```

```

## LE OS RESULTADOS DE UMA VARREDURA DE SCAN PARA ALTAS E BAIXAS INTENSIDADES E
PLOTA OS RESULTADOS ## NO MAPA
#####
out <- read.table("satscan_premio_exp_out_high.col.txt", sep=")
windows(); PlotOnStaticMap(BrasilMap)
## Carrega as funcoes para desenho do CLUSTER PRIMARIO DE ALTA INTENSIDADE
source("circle_builder.R")
circle <- CircleDecimalDegrees(out$V3[1],out$V4[1],radius=out$V5[1],np=25)
PlotOnStaticMap(BrasilMap,circle$lat,circle$lon,add=TRUE,FUN = lines,lwd=2,col='red') ## cex=1.5 ##
Tamanho do ponto
cat('Cluster primario - pvalor: ', out$V10[1], '\n')

## Cluster secundário
if(out$V10[2] < 0.10){
circle <- CircleDecimalDegrees(out$V3[2],out$V4[2],radius=out$V5[2],np=25)
PlotOnStaticMap(BrasilMap,circle$lat,circle$lon,add=TRUE,FUN = lines,lwd=2,col='dark green') ##
cex=1.5 ## Tamanho do ponto
cat('Cluster secundario - pvalor: ', out$V10[2], '\n')
}
## Cluster secundário (3)
if(out$V10[2] < 0.10){
circle <- CircleDecimalDegrees(out$V3[3],out$V4[3],radius=out$V5[3],np=25)
PlotOnStaticMap(BrasilMap,circle$lat,circle$lon,add=TRUE,FUN = lines,lwd=2,col='black') ## cex=1.5 ##
Tamanho do ponto
cat('Cluster secundario (3) - pvalor: ', out$V10[3], '\n')
}
#####
## DESENHA OS CLUSTERS DE BAIXA INTENSIDADE
out <- read.table("satscan_premio_exp_out_low.col.txt", sep=")

circle <- CircleDecimalDegrees(out$V3[1],out$V4[1],radius=out$V5[1],np=25)
PlotOnStaticMap(BrasilMap,circle$lat,circle$lon,add=TRUE,FUN = lines,lwd=2,col='blue') ## cex=1.5 ##
Tamanho do ponto
cat('Cluster primario - pvalor: ', out$V10[1], '\n')

## Cluster secundário
if(out$V10[2] < 0.10){
circle <- CircleDecimalDegrees(out$V3[2],out$V4[2],radius=out$V5[2],np=25)
PlotOnStaticMap(BrasilMap,circle$lat,circle$lon,add=TRUE,FUN = lines,lwd=2,col='blue') ## cex=1.5 ##
Tamanho do ponto
cat('Cluster secundario - pvalor: ', out$V10[2], '\n')
}

```