

# ANÁLISE EXPLORATÓRIA DO CONSUMO PER CAPITA DE ÁGUA RESIDENCIAL E RENDA PER CAPITA, EM UM SETOR DE ABASTECIMENTO DE ÁGUA.

Leandro de Freitas Dadamo  
Instituto Nacional de Pesquisas Espaciais – INPE  
Disciplina: Análises Espaciais de Dados Geográficos

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento das etapas da modelagem estatística espacial, em geral sensível ao tipo de distribuição, à presença de valores extremos e à ausência de estacionariedade.

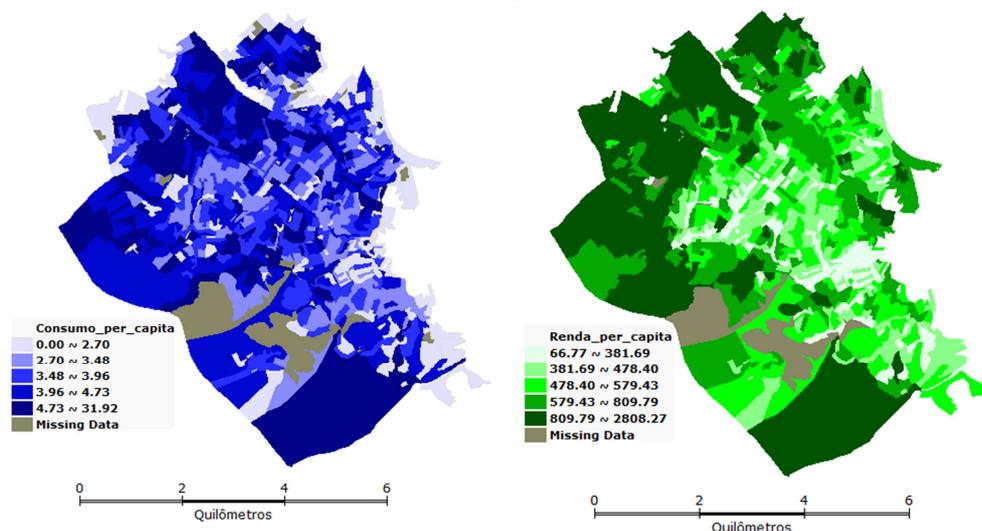
Diante disso, o objetivo deste trabalho, além de contribuir para fixar o conhecimento adquirido nas aulas, é aplicar técnicas de análise exploratória nos dados espaciais de um sistema de abastecimento de água para tentar compreender principalmente o processo de consumo per capita de água residencial e a distribuição de renda por habitante.

No total, foram utilizadas cinco técnicas/métodos de análise exploratória de dados espaciais que serão apresentados, no decorrer do trabalho, os principais conceitos e os respectivos resultados. As técnicas/métodos são: estimador Kernel, índice de Moran Global, Índice de Moran Local – LISA, regressão linear global e regressão geograficamente ponderada – GWR. Os softwares utilizados foram o TerraView e o R.

Os dados utilizados na análise foram renda e população do setor censitário IBGE 2010 e consumo médio de água residencial provenientes da SABESP, entre os meses de agosto a outubro de 2010 (período da coleta de dados do censo demográfico IBGE 2010). Para todas as técnicas utilizadas nesse estudo o consumo de água residencial foi agregado nas áreas do setor censitário, com exceção da técnica do estimador Kernel que utilizou os dados de consumo de água residencial no “suporte ponto”.

Para uma primeira visualização dos dados o Mapa 1 apresenta os dados “consumo per capita de água residencial” (esquerda) e “renda per capita” (direita), utilizando o corte quintil, ou seja, o conjunto de dados ordenados foi dividido em cinco partes iguais.

Mapa 1 - Consumo per capita de água residencial e renda per capita

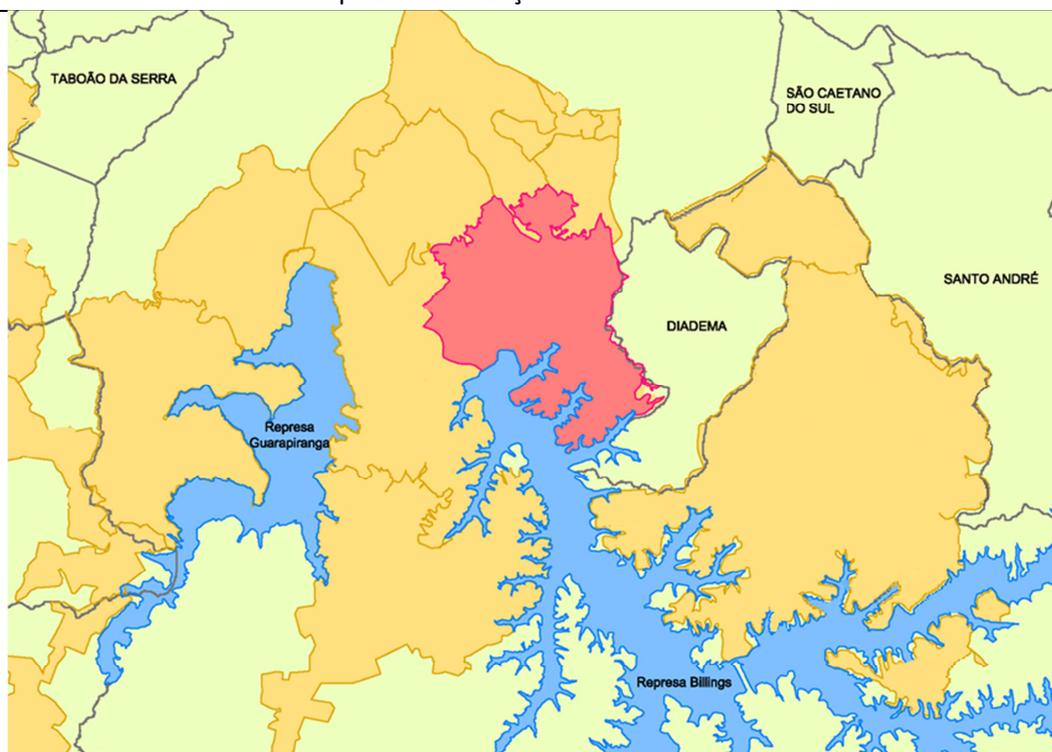


## Área de Estudo

A área de estudo corresponde ao setor de abastecimento de água destacada no Mapa 2 localizada na Zona Sul do município de São Paulo. No tocante da administração política, a área do setor de abastecimento de água está dividida sob a responsabilidade de três Subprefeituras do município de São Paulo. Uma reduzida fração da porção norte do setor de abastecimento se localiza nos limites da Subprefeitura do Jabaquara, a porção noroeste faz parte da área da Subprefeitura de Santo Amaro e a maior parte fica sob a responsabilidade da Subprefeitura de Cidade Ademar.

O setor de abastecimento possui uma área de aproximadamente 33 Km<sup>2</sup> e atende cerca de 520 mil pessoas (IBGE 2010) através de 143.107 ligações residenciais, ou seja, 91,48% do total das ligações. Esta população inserida dentro do setor foi responsável pelo consumo residencial de 22,44 bilhões de litros de água, no ano de 2010.

Mapa 2 – Localização da área de estudo



Para caracterizar socioeconomicamente a área de estudo em questão, é conveniente abordar o Índice Paulista de Vulnerabilidade Social (IPVS 2010, Seade), que é um indicador que visa detalhar as condições de vida da população, com a identificação e a localização espacial das áreas que abrigam os segmentos populacionais mais vulneráveis à pobreza. O IPVS é um indicador resultante da combinação entre duas dimensões – socioeconômica e demográfica –, que classifica cada setor censitário (território contínuo dentro do município que possui em média 300 domicílios) em grupos de vulnerabilidade social.

A Tabela 1 apresenta a quantidade e o percentual de pessoas segregadas em classes do IPVS 2010 para as Subprefeituras que estão inseridas dentro do setor de

abastecimento de água. Percebe-se que a Subprefeitura Cidade Ademar possui uma alta fragilidade, quase 17% da população está localizada em setores censitários classificados com vulnerabilidade muito alta, também conhecido como moradias subnormais ou favelas.

Tabela 1 – População segregada em classes do IPVS 2010

Classe IPVS	Cidade Ademar		Jabaquara		Santo Amaro	
	Pop	%	Pop	%	Pop	%
Não classificado	1.035	0,39	3.204	1,44	250	0,35
Baixíssima vulnerabilidade	4.479	1,68	37.099	16,62	45.252	63,90
Vulnerabilidade muito baixa	82.668	31,01	105.435	47,23	25.320	35,75
Vulnerabilidade baixa	68.050	25,53	38.551	17,27	-	-
Vulnerabilidade média	43.231	16,22	17.483	7,83	-	-
Vulnerabilidade alta (urbanos)	22.265	8,35	2.067	0,93	-	-
Vulnerabilidade muito alta	44.814	16,81	19.382	8,68	-	-
Total	266.542	100,00	223.221	100,00	70.822	100,00

### Estimador de intensidade (Kernel estimation)

Uma alternativa simples para analisar o comportamento de padrões de pontos é a de estimar a intensidade pontual do processo em toda a região de estudo. Para isto, pode-se ajustar uma função bidimensional sobre os eventos considerados, compondo uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área (Bailey e Gatrell, 1995). Esta função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância e pelo peso (atributo) à localização de interesse. Assim temos uma visão geral da intensidade do processo em todas as regiões do mapa.

Este estimador é chamado *kernel estimator* e seus parâmetros básicos são: (a) um raio de influência que define a vizinhança do ponto a ser interpolado e controla o "alisamento" da superfície gerada, um raio muito pequeno irá gerar uma superfície muito descontínua; se for grande demais, a superfície poderá ficar muito amaciada; (b) uma função de estimação com propriedades de suavização do fenômeno.

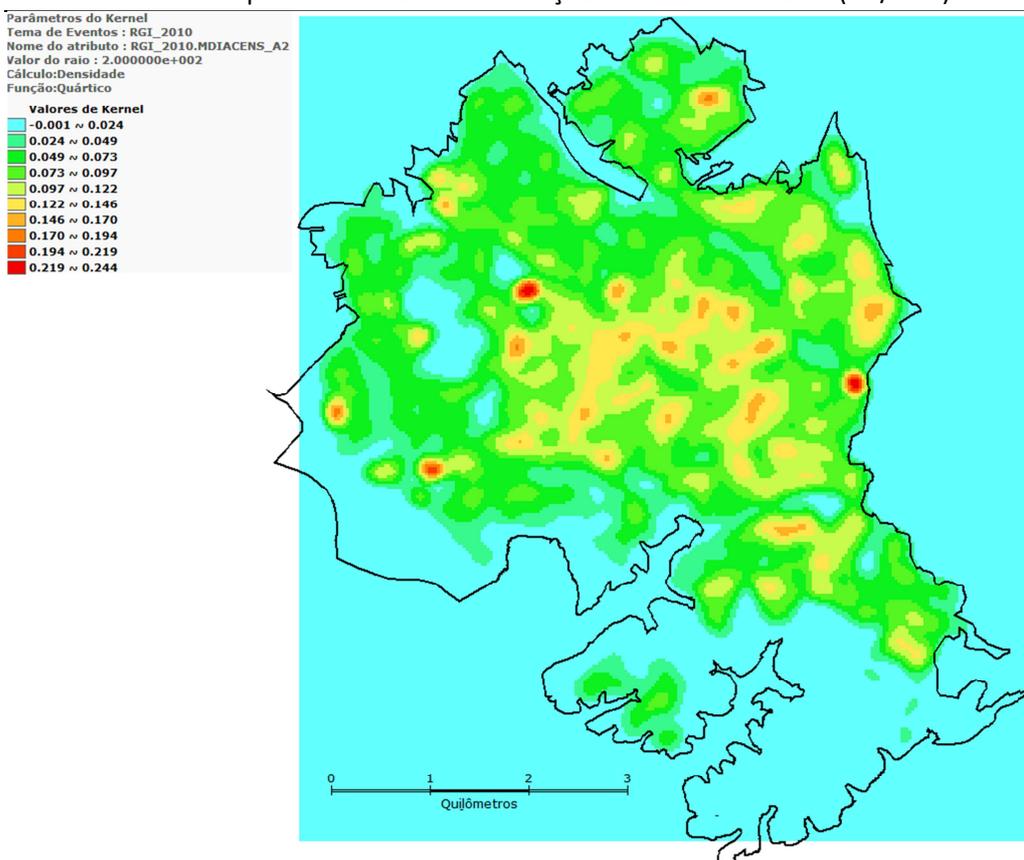
Tendo em vista que a base cadastral da SABESP é georreferenciada, foi possível aplicar esse conceito utilizando o software TerraView que é um aplicativo geográfico, um SIG (Sistema de Informação Geográfica), visualizador de dados geográficos com recursos avançados de consulta e análise geográfica destes dados.

O evento espacial observado foi o RGI, ou hidrômetro, considerando o respectivo consumo médio mensal em m<sup>3</sup>, para o período de agosto de 2010 a outubro de 2010 (período da coleta de dados do censo demográfico IBGE 2010) e o raio utilizado para a elaboração do mapa foi de 200 metros. A aplicação do estimador de intensidade gerou uma superfície raster que pode ser visualizada no Mapa 3.

As cores quentes (vermelho) representam uma região com consumo maior em relação às cores frias (azul). A superfície interpolada revela que a região central possui maior consumo. As pequenas regiões vermelhas espalhadas no mapa estão associadas à

instalação de condomínios verticais, onde todo o consumo de água (geralmente alto) é lido por um único hidrômetro e representado eu um ponto.

Mapa 3 – Padrão de distribuição do consumo médio (m<sup>3</sup>/mês)



### Índice de Moran Global

O índice de Moran global visa identificar a estrutura de correlação espacial que melhor descreve os dados. A ideia básica é a caracterização da dependência espacial, mostrando como os valores estão correlacionados no espaço. De uma forma geral, a autocorrelação espacial mede o quanto o valor observado de um atributo numa região é independente dos valores desta mesma variável nas localizações vizinhas. Valores positivos indicam para correlação direta, negativos indicam correlação inversa e valores nulos significam independência espacial. Uma vez calculado, é importante estabelecer sua validade estatística. Os índices globais das variáveis de renda e consumo de água residencial total e per capita estão apresentados na Tabela 2.

Tabela 2 – Índice de Moran Global e p-valor

	Consumo per capita de água residencial	Renda per capita
Valor do Índice	0,0837	0,6220
p-valor	0,0100	0,0100

Para a variável de consumo per capita de água residencial verificou-se que não há presença de autocorrelação espacial, uma vez que o índice de Moran teve valor de 0,0837. Já para a variável renda per capita foi possível afirmar que há evidências de

autocorrelação espacial, pois o índice de Moran foi de 0,6220. Analisando o Mapa 1 apresentado anteriormente, já era possível perceber de forma intuitiva a presença de autocorrelação espacial para a variável renda per capita, entretanto o uso de *percentis* pode mascarar diferenças significativas em valores extremos e dificultar a identificação de áreas críticas.

### **Índice de Moran Local - LISA**

Os indicadores globais de autocorrelação espacial, como o Índice de Moran global, fornecem um único valor como medida da associação espacial para todo o conjunto de dados de área, que é útil para caracterização de toda a região de estudo. Por contraste, muitas vezes é desejável examinar padrões numa escala de maior detalhe, para verificar se a hipótese de estacionariedade do processo verifica-se localmente.

Para tanto, é preciso utilizar indicadores de associação espacial que possam ser associados às diferentes localizações de uma variável distribuída espacialmente. Essa metodologia utiliza o Índice de Moran Local para encontrar a correlação espacial dessas áreas. Por se tratar de um indicador local, tem-se um valor específico de correlação para cada área o que acarreta na identificação de “clusters” (objetos com valores de atributos semelhantes), “outliers” (objetos anômalos) e a presença de mais de um regime espacial.

Segundo Anselin (1995), um indicador local de associação espacial (LISA) tem que atender a dois objetivos: (a) permitir a identificação de padrões de associação espacial significativo e (b) ser uma decomposição do índice global de associação espacial.

Através da ferramenta “estatística espacial” do software TerraView, foi possível calcular os indicadores locais, o Diagrama de Espalhamento de Moran e gerar mapas de forma simples e rápida.

O diagrama de espalhamento de Moran para as variáveis de consumo per capita de água residencial e renda per capita é apresentado a seguir, nos Diagrama 1 e Diagrama 2, respectivamente.

Os pontos localizados nos quadrantes Q1 e Q2 indicam regiões de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes, já os pontos localizados nos quadrantes Q3 e Q4 indicam regiões de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos. A reta bissetriz que passa pelos quadrantes Q1 e Q2 representa a maior associação espacial entre a área analisada e seus vizinhos, sendo assim o Diagrama 2 está mais próxima dessa tendência do que o Diagrama 1.

Diagrama 1 – Diagrama de Espalhamento de Moran – Consumo per capita de água residencial

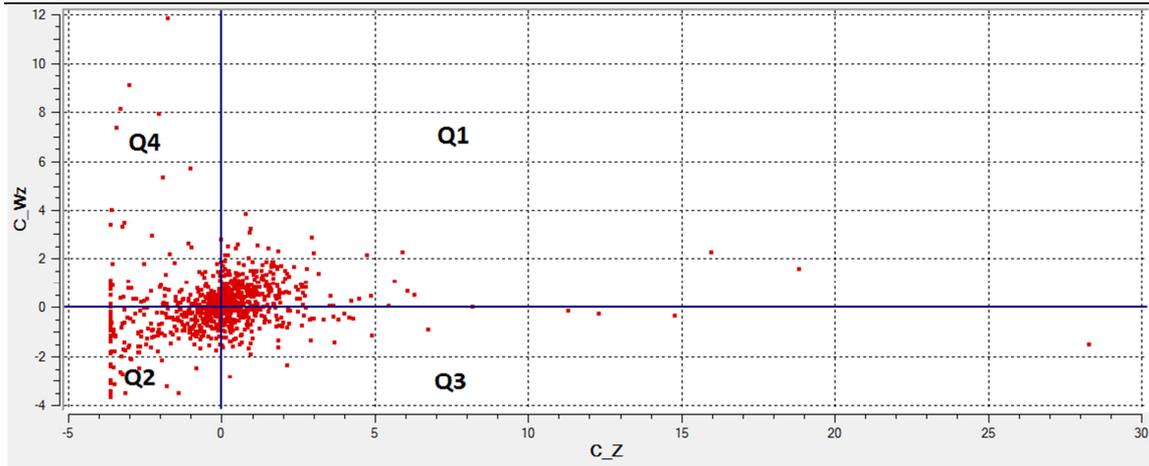
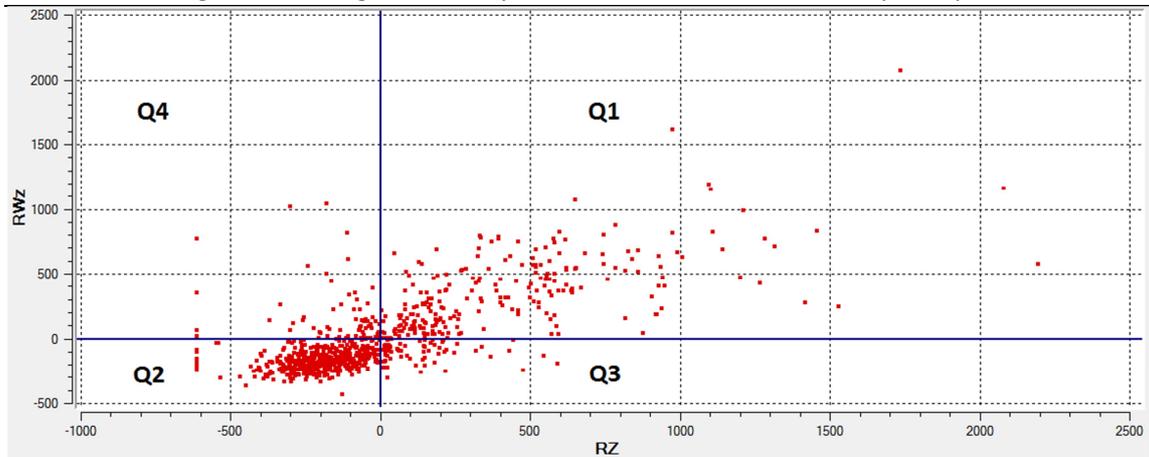


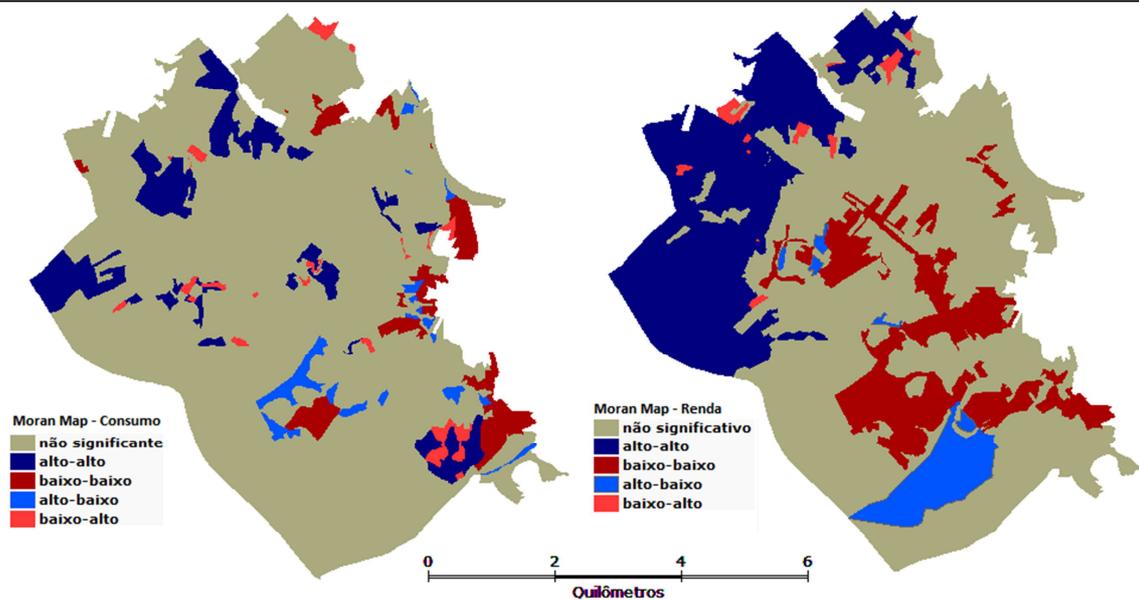
Diagrama 2 - Diagrama de Espalhamento de Moran – Renda per capita



O diagrama de espalhamento de Moran pode ser apresentado na forma de um mapa coroplético bidimensional, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento, onde Q1 é representado como alto-alto, Q2=baixo-baixo, Q3=alto-baixo e Q4=baixo-alto.

O Moran map, Mapa 4, apresenta somente as regiões para os quais os valores do índice local são significantes (com intervalo > 95%). As regiões de tonalidades cinza não possui significância satisfatória, sendo que 84% dos setores censitários apresentaram um índice insignificante para a variável de consumo per capita de água residencial e 65% para a variável renda per capita. Este último apresentou evidências de autocorrelação espacial na região central e no sudeste da área de estudo (baixo-baixo) onde a renda per capita é menor. Para a região noroeste (alto-alto), porção localiza na subprefeitura de Santo Amaro, a renda per capita é mais elevada.

Mapa 4 – Moran Map – Consumo per capita de água residencial (esquerda) e renda per capita (direita)



### Regressão Linear Global

Inicialmente, uma maneira de estimar a relação entre as variáveis é através de um modelo de regressão linear global (NETER et al., 1996). Sendo assim, todas as observações consideradas na análise são utilizadas para estimar um único modelo:

$$CONSUMO_i = \beta_0 + \beta_1 \cdot RENDA_i + \varepsilon$$

onde:  $CONSUMO_i$  representa, em  $m^3/mês$ , o consumo médio mensal per capita de água residencial no setor censitário  $i$  (variável resposta),  $RENDA_i$  corresponde à renda mensal per capita no setor censitário  $i$  (variável explicativa),  $\beta_0$  e  $\beta_1$  são os parâmetros a serem estimados, e  $\varepsilon$  representa o erro.

O Quadro 1 apresenta os resultados do modelo e pode-se observar uma associação positiva ( $\beta = 1,035 \times 10^{-3}$ ) e estatisticamente significativa entre o consumo de água per capita e a renda per capita. Portanto, de acordo com este modelo, o incremento de uma unidade (R\$ 1,00) na renda média per capita da população está associado a um incremento de  $1,035 \times 10^{-3} m^3/mês$  (ou 1,035 l/mês) no consumo de água residencial por habitante.

Em relação à qualidade do ajuste, foi possível constatar através do coeficiente de determinação ( $R^2$ ) que o modelo estimado é capaz de descrever apenas 2,6% da variabilidade observada na variável consumo per capita de água.

---

### Quadro 1 – Resumo do modelo linear global – Software R

---

```
Call:
lm(formula = C_pc_c_1 ~ R_pc_1, data = shp.base)

Residuals:
    Min       1Q   Median       3Q      Max
-5.3606 -0.7642  0.0427  0.7125 28.3046

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.0224340   0.1535475   19.684 < 2e-16 ***
R_pc_1       0.0010353   0.0002157    4.799 1.88e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.237 on 853 degrees of freedom
Multiple R-squared:  0.02629,    Adjusted R-squared:  0.02515
F-statistic: 23.03 on 1 and 853 DF,  p-value: 1.88e-06
```

---

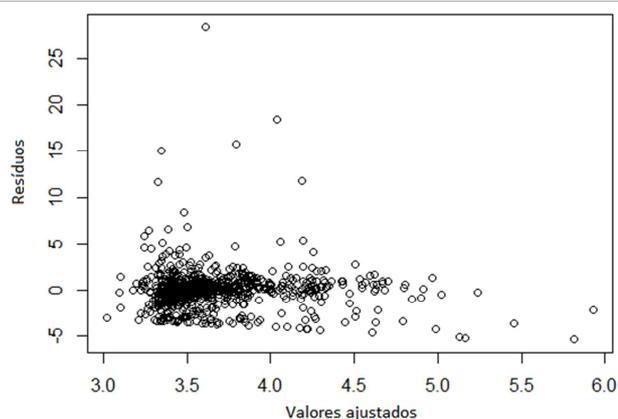
O valor do coeficiente de determinação próximo a zero sugere fraca relação entre as variáveis. Entretanto, é importante destacar que o modelo representa de maneira constante a associação entre consumo e renda per capita, gerando resultados globais que se assumem válidos para todos os setores censitários da área de estudo e, também, em uma análise preliminar dos valores dos resíduos, mostrados no Gráfico 1, é possível perceber alguns problemas, como “outliers” e possível heteroscedasticidade.

Para o diagnóstico de heteroscedasticidade, tenta-se encontrar alguma tendência no gráfico dos valores ajustados versus resíduos. Por isso, se os pontos estão aleatoriamente distribuídos em torno do 0, sem nenhum comportamento ou tendência, tem-se indícios de que a variância dos resíduos é homoscedástica. Como o Gráfico 1 mostra uma distribuição não aleatória em torno do 0 é um indicativo da presença de heteroscedasticidade.

---

Gráfico 1 – Valores ajustados versus resíduos

---



### Regressão Geograficamente Ponderada – GWR

Como mencionado anteriormente, o modelo de regressão simples representa de maneira constante a associação entre consumo per capita de água residencial e renda per capita para toda a área de estudo, além disso, se considerar a heterogeneidade dos processos socioeconômicos vigentes em um país como o Brasil, é bastante provável que a relação entre as variáveis analisadas ocorra de maneira distinta. Neste

caso, dada sua incapacidade de incorporar variações espaciais, um modelo tradicional de regressão apresentaria um poder explicativo reduzido.

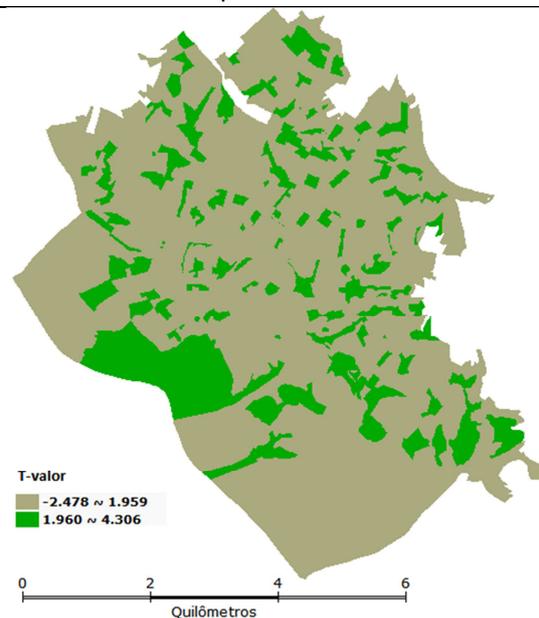
O modelo de regressão geograficamente ponderado (GWR – Geographically Weighted Regression, FORTHERINGHAM et al., 2002) é uma técnica espacial de análise exploratória que trabalha com indicadores locais ao invés dos modelos globais, mais tradicionais. De forma geral, o GWR ajusta um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto. Neste estudo, para a determinação dessas distâncias o GWR utilizou uma função *kernel* com largura de banda adaptativa. Após a definição da largura da banda mais apropriada, os coeficientes locais da regressão, erros,  $R^2$  podem ser obtidos e mapeados.

O software utilizado para executar o GWR foi o R, através do pacote “SPGWR”. Analisando os resultados do modelo foi constatada uma associação positiva e significativa entre o consumo per capita de água residencial e renda per capita em apenas 21% dos modelos estimados (t-valor associado ao  $\beta_1 > 1,96$ ), o Mapa 5 apresenta os setores censitários cujo coeficiente estimado foi estatisticamente significativo. De acordo com o modelo GWR, também é possível concluir que o incremento de uma unidade (R\$ 1,00) na renda média per capita da população está associado a um incremento de exatamente 1,0 l/mês no consumo de água residencial por habitante.

---

Mapa 5 – T-valor

---

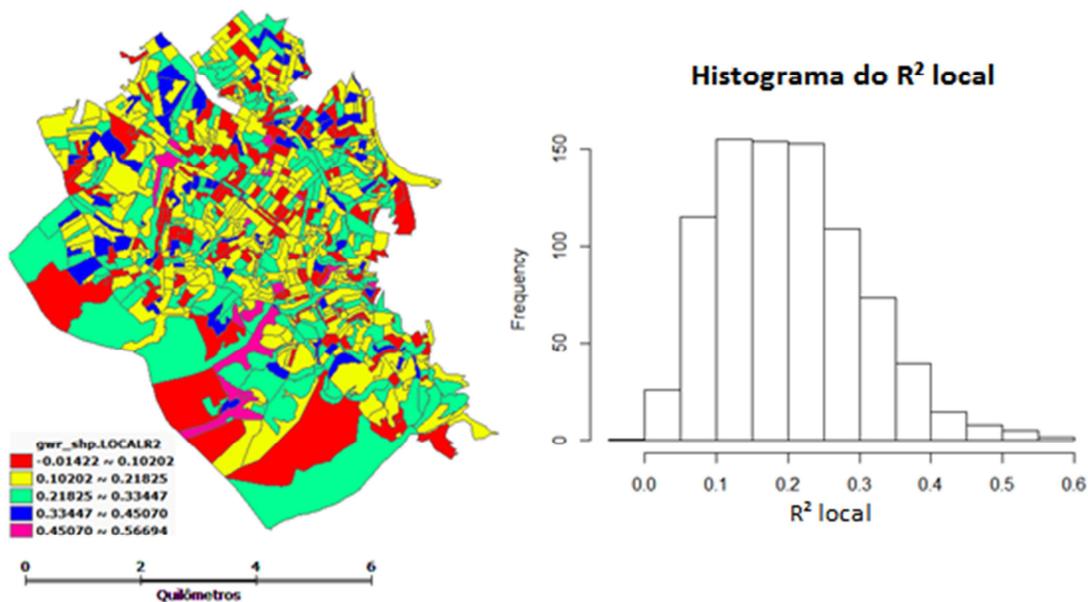


A função “SPGWR” gera um objeto espacial de saída com um quadro de dados (Spatial Data Frame – SDF) contendo todos os parâmetros para a execução do modelo, sendo assim é possível mapear todos ou, pelo menos, os parâmetros mais importantes para auxiliar na interpretação do modelo GWR. Além do parâmetro t-valor, é interessante gerar um mapa e analisar visualmente o coeficiente local estimado para a variável

renda per capita ( $\beta_1$  ou inclinação na reta), e o coeficiente de determinação local ( $R^2$ ) para cada setor censitário.

Comparando a qualidade do ajuste ( $R^2$ ) do modelo GWR com a regressão linear simples observa-se uma melhora na qualidade do ajuste que conseguiu explicar 19,40 % da variabilidade observada na variável "consumo per capita de água residencial", entretanto ainda é um valor baixo. A Figura 1 apresentada a seguir mostra um mapa com os coeficientes locais de determinação ( $R^2$ ) calculados para todos os setores censitários juntamente com o histograma mais para ilustrar do que para concluir algo, pois 79% dos coeficientes locais não possui significância estatística e qualquer conclusão poderia estar equivocada.

Figura 1 – Mapeamento e histograma do coeficiente de determinação local ( $R^2$ )



## Conclusão

Após aplicar cinco técnicas de análise exploratórias pode-se concluir que o conjunto de dados precisa ser tratado preliminarmente para obter maior grau de significância nas análises e conseqüentemente conclusões com maiores certezas.

Com exceção da função kernel, que é possível perceber de forma precisa a intensidade pontual do processo de consumo de água residencial, verificou-se nas outras técnicas de análise exploratória de dados geográficos problemas como, "outliers", heteroscedasticidade na região de estudo e pouco significância nos coeficientes locais.

Portanto, é muito difícil e arriscado realizar qualquer tipo de análise em um conjunto de dados dos quais somente um quinto possui significância estatística, segundo o modelo GWR, e qualquer conclusão a partir destas análises estará sujeita a erros.

## Referências Bibliográficas

ANSELIN, L. Local indicators of spatial association – LISA. *Geographical Analysis*. 27, 93-115. 1995.

Bailey, T.; Gatrell, A. "Interactive Spatial Data Analysis". London, Longman Scientific and Technical, 1995

FEITOSA, Flávia da Fonseca et al. POPULAÇÃO E CONSUMO URBANO DE ÁGUA NO BRASIL: INTERFACES E DESAFIOS. Simpósio Brasileiro de Recursos Hídricos, 2013.

FOTHERINGHAM, A. S.; BRUNDSON, C.; CHARLTON, M. (2002). Geographically weighted regression: the analysis of spatially varying relationships. Chichester: J. Wiley & Sons. 269 p.

IBGE – Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010.

NETER, J. et al. (1996) Applied Linear Statistical Models. Chicago: Irwin. 1396 p.

R Development Core Team (2008) R Foundation for Statistical Computing, Vienna, Austria.

Suzana Fucks; Marília Sá Carvalho; Gilberto Câmara; Antonio Miguel V. Monteiro; Análise Espacial de Dados Geográficos. São José dos Campos, INPE, 2003 - on-line (3a. edição, revista e ampliada). Versão em papel - Editora: EMBRAPA Cerrados. Previsão: Dezembro 2004.

TerraView 4.1.0. São José dos Campos, SP: INPE, 2010. Disponível em: [www.dpi.inpe.br/terraview](http://www.dpi.inpe.br/terraview). Acesso em: 01/10/2013.