

## ANÁLISE ESTATÍSTICA ESPACIAL DO CÂNCER NOS MUNICÍPIOS PAULISTAS<sup>1</sup>

Ana Cristina Maurício Ferreira<sup>2</sup>, Delton Campanhã de Moraes<sup>3</sup>, Luis Gustavo Leite Marques<sup>3</sup>

<sup>1</sup> Departamento de Sistemas Biomédicos – FATEC/BAURU

<sup>2</sup> Departamento de Computação – FATEC/BAURU

<sup>3</sup> Departamento de Sistemas Biomédicos – FATEC/BAURU

**ABSTRACT.** *Cancer present as a major cause of illness and death in Brazil, where, since 2003, approximately 17% of deaths with known cause, notified in 2007 Mortality Information System. Thus, considering cancer as a phenomenon occurring spatialized aimed to study the spatial correlation of cancer mortality occurred in the municipalities of São Paulo, or verifies that the occurrences of a given municipality can influence events in neighboring municipalities. We used a base map of the State of São Paulo, containing a vector representation of the municipal limits, provided by the National Institute for Space Research (INPE). Regarding the data regarding the incidence of cancer mortality in the counties, these refer to the year 2010 and were collected from the Database of the Unified Health System (DATASUS). For data analysis, we used spatial statistical tools, which aims to verify the existence of spatial autocorrelation of cancer mortality rates. We used spatial statistics tools, specifically the Global Index and Local Moran. As results were found four regions in which the spatial autocorrelation is more evident, those regions that involve groups of municipalities.*

**KEYWORDS:** *cancer, GIS, spatial analysis, spatial statistics.*

**RESUMO.** *O câncer apresenta-se como importante causa de doença e morte no Brasil, constituindo, desde 2003, aproximadamente 17% dos óbitos por causa conhecida, notificados em 2007 no Sistema de Informações sobre Mortalidade. Assim, considerando o câncer como um fenômeno de ocorrência espacializada, objetivou-se estudar a correlação espacial da mortalidade por câncer ocorrida nos municípios do Estado de São Paulo, ou seja, verificar se as ocorrências de um determinado município podem influenciar as ocorrências dos municípios vizinhos. Foi utilizada uma base cartográfica do Estado de São Paulo, contendo a representação vetorial dos limites municipais, disponibilizada pelo Instituto Nacional de Pesquisas Espaciais (INPE). Em relação aos dados referentes à ocorrência de mortalidade por câncer nos municípios paulistas, estes se referem ao ano de 2010 e foram coletados junto ao Banco de Dados do Sistema Único de Saúde (DATASUS). Para a análise de dados, foram utilizadas ferramentas de estatística espacial, que tem por finalidade verificar a existência de autocorrelação espacial das taxas de mortalidade por câncer. Foram utilizados ferramentas de estatística espacial, especificamente o Índice Global e Local de Moran. Como resultados foram encontradas quatro regiões em que a autocorrelação espacial é mais evidente, regiões estas que envolvem agrupamentos de municípios.*

**PALAVRAS-CHAVE:** *câncer, geoprocessamento, análise espacial, estatística espacial.*

<sup>1</sup> Trabalho apresentado na Feira Tecnológica do Centro Paula Souza, 2011.

## 1. Introdução

A saúde é um assunto fundamental na sociedade desde a antiguidade, sendo atualmente considerada, através da expectativa de vida, como um dos três pilares para o cálculo do Índice de Desenvolvimento Humano (IDH), desenvolvido pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), da Organização das Nações Unidas (ONU). Dentre as enfermidades que mais afetam a população atualmente estão as neoplasias – o câncer – e suas diversas formas de manifestação.

Este trabalho propõe-se a verificar a existência de uma possível autocorrelação espacial na existência de casos de neoplasias nos municípios do estado de São Paulo, através do uso do geoprocessamento, que possibilita a realização de análises espaciais e também a visualização destas análises através de mapas temáticos. O geoprocessamento coloca-se, nesse sentido, como uma ferramenta de importância fundamental, permitindo a análise geoestatística dos municípios e de sua vizinhança.

O tema deste trabalho foi escolhido devido à busca pela confirmação ou negação da hipótese de que os casos de câncer podem estar relacionados a algum padrão ou característica espacial local específica, o que dá suporte para novos trabalhos na área, que venham a analisar essas regiões especificamente para identificar esse possível padrão ou característica comum. É possível, com essa investigação, buscar quais fatores contribuem para as ocorrências nestes locais em específico. O objetivo geral desse trabalho, portanto, foi verificar a existência de correlação espacial das taxas de mortalidade por neoplasias entre os municípios do Estado de São Paulo.

## 2. Revisão De Literatura

A revisão de literatura busca trazer informações sobre os principais temas relacionados a este trabalho, a começar pelo objeto de estudo, que é o câncer. Mostra também os principais conceitos de geoprocessamento, análise e estatística espacial.

### 2.1 Câncer

Segundo o INCA (2011), câncer é o nome dado a um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos. Dividindo-se rapidamente, estas células tendem a ser muito agressivas e incontroláveis, determinando a formação de tumores malignos.

Tumor é um termo genérico que indica um aumento anormal de uma parte ou da totalidade de um tecido, podendo ser benigno ou maligno. A diferença básica entre o tumor benigno e maligno está na velocidade de multiplicação das células com aumento do tumor e sua capacidade de gerar metástases pelo corpo, isto é, espalhar para outras partes do corpo. Pode-se dizer que no tumor benigno, as células ficam envolvidas por uma membrana, que impede que elas se desenvolvam e se espalhem; já no maligno, pode começar no estômago e espalhar-se por todo o sistema digestivo ou qualquer outra parte do corpo, como fígado, pulmão e ossos, por exemplo. Desta forma, o câncer pode causar invasão destrutiva de órgãos normais por estas células, por extensão direta ou por disseminação à distância, que pode ser através do sangue, linfa ou superfície serosa (SASSE, 2002).

De acordo com o Hospital Erasto Gaertner (2011), existem estruturas responsáveis pelo crescimento celular (proto-oncogenes), assim como existem estruturas responsáveis pela morte celular (oncogenes supressores). Tais estruturas desempenham importante papel na regulação do

ciclo celular. De certa forma, um descontrole entre essas estruturas, tanto para mais, quanto para menos, seria responsável pelo surgimento do câncer.

Dentre as diversas causas responsáveis por esse descontrole, ou seja, pelo surgimento do câncer, encontram-se os carcinógenos ambientais (cigarro, álcool, radiação ionizantes (luz solar), produtos químicos, má alimentação etc.); os fatores endógenos (alterações hormonais, obesidade, envelhecimento, etc.) e os fatores genéticos.

Segundo Sasse (2002), o câncer é fundamentalmente uma doença genética. Quando o processo neoplásico se instala, a célula-mãe transmite às células filhas a característica neoplásica. Isso quer dizer que, no início de todo o processo está uma alteração no DNA de uma célula. Esta alteração inicial recebe o nome de estágio de iniciação, porém, uma só alteração no DNA não causa câncer. São necessárias várias alterações em seqüência para que essa célula torne-se cancerosa. O estágio de promoção é o segundo estágio da carcinogênese; nele, as células geneticamente alteradas, ou seja, "iniciadas", sofrem o efeito dos agentes. A célula iniciada é transformada em célula maligna, de forma lenta e gradual. Para que ocorra essa transformação, é necessário um longo e continuado contato com o agente cancerígeno promotor. O estágio de progressão é o terceiro e último estágio e se caracteriza pela multiplicação descontrolada e irreversível das células alteradas. Nesse estágio o câncer já está instalado, evoluindo até o surgimento das primeiras manifestações clínicas da doença.

Segundo o INCA (2011), o câncer pode surgir em qualquer parte do corpo, mas alguns órgãos são mais afetados do que outros. Entre os mais afetados estão pulmão, mama, colo do útero, próstata, cólon e reto (intestino grosso), pele, estômago, esôfago, medula óssea (leucemias) e cavidade oral (boca). Cada órgão, por sua vez, pode ser afetado por tipos diferenciados de tumor, menos ou mais agressivos.

7

## 2.2 O Estado de São Paulo

São Paulo é uma das 27 unidades federativas do Brasil e está localizado no sul da região Sudeste, tendo como limites os estados de Minas Gerais, Rio de Janeiro, Paraná e Mato Grosso do Sul, além do Oceano Atlântico, a leste. O estado ocupa uma área de 248.808,8 quilômetros quadrados e sua capital é a cidade de São Paulo. Tem a maior população do Brasil, totalizando mais de 40 milhões de habitantes (dos quais cerca de três milhões de imigrantes, de 70 diferentes nacionalidades) distribuídos em 645 municípios. O estado é a mais rica das unidades federativas e também figura entre os estados com alto IDH. Responsável por mais de 31% do PIB do país, São Paulo é considerado o "motor econômico" do Brasil por possuir melhor infraestrutura, mão de obra qualificada, fabricar produtos de alta tecnologia, além de abrigar o maior parque industrial e a maior produção econômica do país (SEADE, 2011).

A Secretaria de Estado da Saúde de São Paulo (SES-SP) é responsável pela formulação da Política Estadual de Saúde e de suas diretrizes, norteadas pelos princípios do Sistema Único de Saúde - SUS, que tem como propósitos promover a saúde priorizando as ações preventivas, democratizando as informações relevantes para que a população conheça seus direitos e os riscos à sua saúde.

## 2.3 Geoprocessamento

Segundo Câmara *et al.* (2001), as primeiras tentativas de automatizar parte do processamento de dados com características espaciais aconteceram na Inglaterra e nos Estados Unidos, nos anos 50, com o objetivo principal de reduzir os custos de produção e manutenção de mapas, mas esses sistemas ainda não eram chamados (e nem podem ser considerados) 'sistemas de informações

geográficas', pois as aplicações, os dados e os sistemas ainda eram muito precários. As pesquisas nessa época envolveram a botânica, na Inglaterra, e o volume de tráfego, nos Estados Unidos. Na década de 60 surgiram os primeiros Sistemas de Informação Geográfica (SIG), especialmente no Canadá, como parte de um programa do governo voltado para os recursos naturais. A partir dos anos 70, com o desenvolvimento da informática é que começa a se desenvolver comercialmente o uso desses sistemas, surgem diversos conceitos e produtos aplicados aos SIGs e à cartografia, e também a expressão *Geographic Information System* (GIS). Finalmente nos anos 80 começa o acelerado desenvolvimento e a massificação dos SIGs associada à evolução dos meios de comunicação, em especial da informática. Com a criação do NCGIA - *National Centre for Geographical Information and Analysis* em 1989, nos Estados Unidos, o Geoprocessamento é estabelecido como disciplina científica independente.

A introdução do Geoprocessamento no Brasil inicia-se a partir do esforço de divulgação e formação de pessoal feito pelo prof. Jorge Xavier da Silva, da Universidade Federal do Rio de Janeiro (UFRJ), no início dos anos 80. A vinda ao Brasil, em 1982, do Dr. Roger Tomlinson, responsável pela criação do primeiro SIG (o *Canadian Geographical Information System*), incentivou o aparecimento de vários grupos interessados em desenvolver a tecnologia (CÂMARA *et al.*, 2001). Dentre esses grupos estão a UFRJ, pioneira nesse tipo de estudos no Brasil e o INPE, com sede em São José dos Campos, interior de São Paulo, fundado em 1971. De acordo com Almeida (2010), já na década de 1970, os projetos do INPE eram voltados para estudos que se utilizavam de dados de satélites de observação da Terra, meteorológicos e de comunicação. Atualmente, o INPE possui sedes em várias cidades do Brasil, assumindo também uma posição de ponta em termos de ciência e aplicações de tecnologias espaciais e contando com um moderno centro de rastreamento e controle de satélites.

Também é de autoria do INPE o SPRING, um SIG com funções de processamento de imagens, análise espacial, modelagem numérica de terreno e consulta a bancos de dados espaciais (CÂMARA *et al.*, 1996), e referência brasileira no Geoprocessamento.

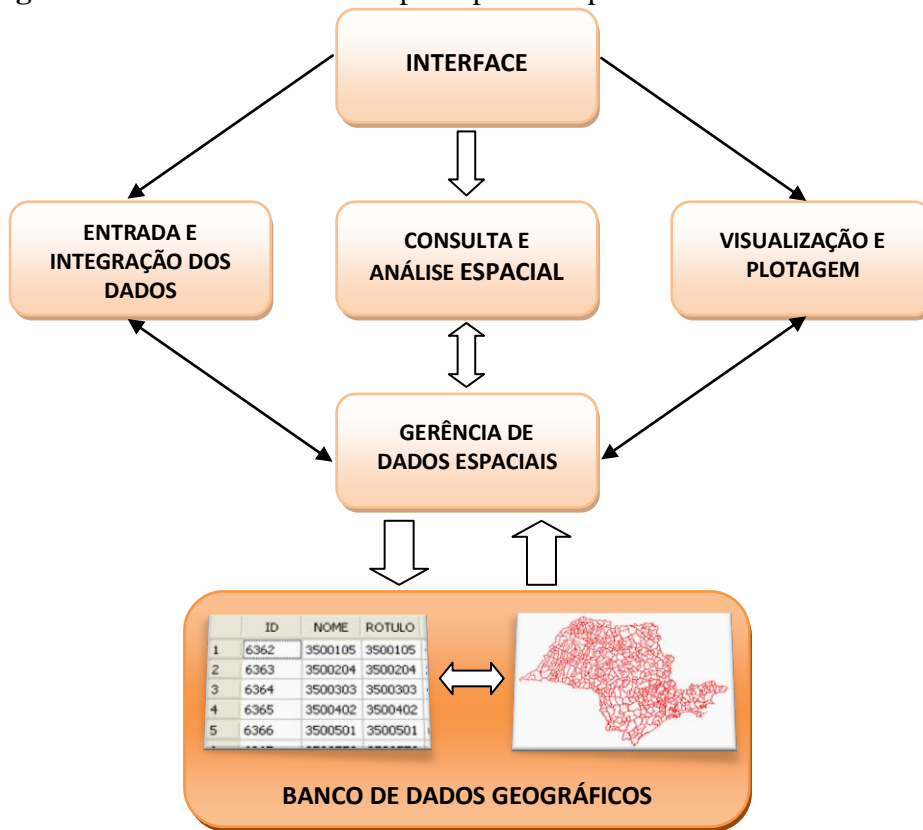
De acordo com Câmara *et al.* (2001), numa visão abrangente, pode-se indicar que um SIG tem os seguintes componentes:

- Interface com usuário;
- Entrada e integração de dados;
- Funções de consulta e análise espacial;
- Visualização e plotagem;
- Armazenamento e recuperação de dados (organizados sob a forma de um banco de dados geográficos).

Estes componentes se relacionam de forma hierárquica, distinguindo-se em três níveis. No nível mais próximo do usuário, a interface homem-máquina define como o sistema é operado e controlado. No nível intermediário, um SIG apresenta mecanismos de processamento de dados espaciais (entrada, edição, análise, visualização e saída). Por fim, no nível mais interno do sistema, um sistema gerencia os bancos de dados geográficos oferecendo o armazenamento e a recuperação de seus dados espaciais associados a seus atributos (Figura 1).

A relação entre os dados geográficos e as funções de processamento do SIG é feita por mecanismos de seleção e consulta que definem restrições sobre o conjunto de dados. A partir dessa seleção e consulta um resultado que obedece a essas restrições é obtido, identificando as características geográficas e espaciais desejadas.

**Figura 1** - Relacionamento dos principais componentes ou subsistemas de um SIG



Fonte: Adaptado de CÂMARA *et al.*, 2001

Cada Sistema de Informações Geográficas, em função de seus objetivos e necessidades, implementa estes componentes de forma distinta, mas todos os subsistemas citados devem estar presentes num SIG.

## 2.4 Análise e Estatística Espaciais

Análise espacial é um procedimento de estudo que utiliza ferramentas de geoestatística, procurando analisar padrões espaciais e verificar a aleatoriedade destes padrões. Permitem descrever e analisar a distribuição espacial de uma determinada variável (no caso a taxa de mortalidade por câncer), identificar situações atípicas, não só em relação à área em questão, mas também em relação aos vizinhos e buscar a existência de padrões na dependência espacial presentes no fenômeno. O desafio é medir o grau de associação espacial entre áreas de uma ou mais variáveis (SERRANO E VALCARCE, 2000).

A análise espacial está fundamentada na primeira lei da geografia: “todas as coisas se parecem, mas coisas mais próximas são mais parecidas que as mais distantes”. É necessário modelar a distribuição do parentesco entre os dados, por meio de técnicas de estatística espacial, para que se possa ter maior grau de confiabilidade nos dados e no entendimento dos problemas estudados. A informação que se busca através do cálculo de autocorrelação espacial é de quanto o valor da taxa de mortalidade por neoplasias é parecido com o valor dessa taxa no município mais próximo e de

quanto é diferente no município mais distante. A autocorrelação espacial ocorre quando observações organizadas no espaço influenciam-se mutuamente (FERREIRA e RAIA JUNIOR, 2010).

Várias são as ferramentas de estatística espacial e, neste trabalho foram utilizados os Índices Global e Local de *Moran*. O cálculo **Índice Global de Moran**, fornece uma medida geral da associação espacial existente no conjunto dos dados, variando de [-1, 1]. Dados com baixa associação espacial resultam em um índice baixo, próximo de zero; valores positivos e negativos significam autocorrelação espacial positiva e negativa, respectivamente. O Índice Global de *Moran* ( $I$ ) é calculado através da expressão (1), em que  $n$  é o número de municípios do Estado;  $Z_i$  é a diferença entre o valor do atributo (mortalidade por neoplasia) do município  $i$  e a média dos atributos dos demais municípios;  $Z_j$  é a diferença entre o valor do atributo dos vizinhos do município  $i$  e a média de todos os atributos dos demais municípios;  $w_{ij}$  são os pesos atribuídos segundo a relação topológica entre os municípios  $i$  e  $j$ .

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} Z_i Z_j}{\sum_{i=1}^n Z_i^2} \quad (1)$$

O **Índice Local de Moran**, tem por finalidade verificar a existência de padrões locais de associação espacial. Os indicadores locais produzem um valor específico para cada objeto, permitindo assim, a identificação de agrupamentos de objetos com valores de atributos semelhantes (*clusters*), objetos anômalos (*outliers*) ou de mais de um regime espacial. O Índice Local de *Moran* é calculado segundo a expressão (2), em que  $W_{ij}$  é o valor na matriz de vizinhança do município  $i$  com o município  $j$  em função da distância, e  $Z_i$  e  $Z_j$  são desvios em relação à média.

$$I = Z_i \sum w_j Z_j \quad (2)$$

### 3. Materiais e Métodos

Para o desenvolvimento deste projeto foram utilizados os materiais e métodos descritos a seguir.

11

#### 3.1 Programa

Para a análise geoestatística e confecção dos mapas do trabalho foi utilizado o programas SPRING na versão 5.1.3, do INPE – Instituto Nacional de Pesquisas Espaciais.

### 3.2 Documentos

Os documentos utilizados foram:

- Malha de municípios paulistas, do censo demográfico de 2000, do IBGE através do “Banco de Dados São Paulo” disponibilizado pelo INPE.
- Dados de mortalidade (número absoluto) por neoplasias, por município, no estado de São Paulo obtidos no sistema de saúde pública do estado de São Paulo (baseado no DATA-SUS), com posterior cálculo da Taxa de Mortalidade por Neoplasia - (população/mortalidade)\*100000.

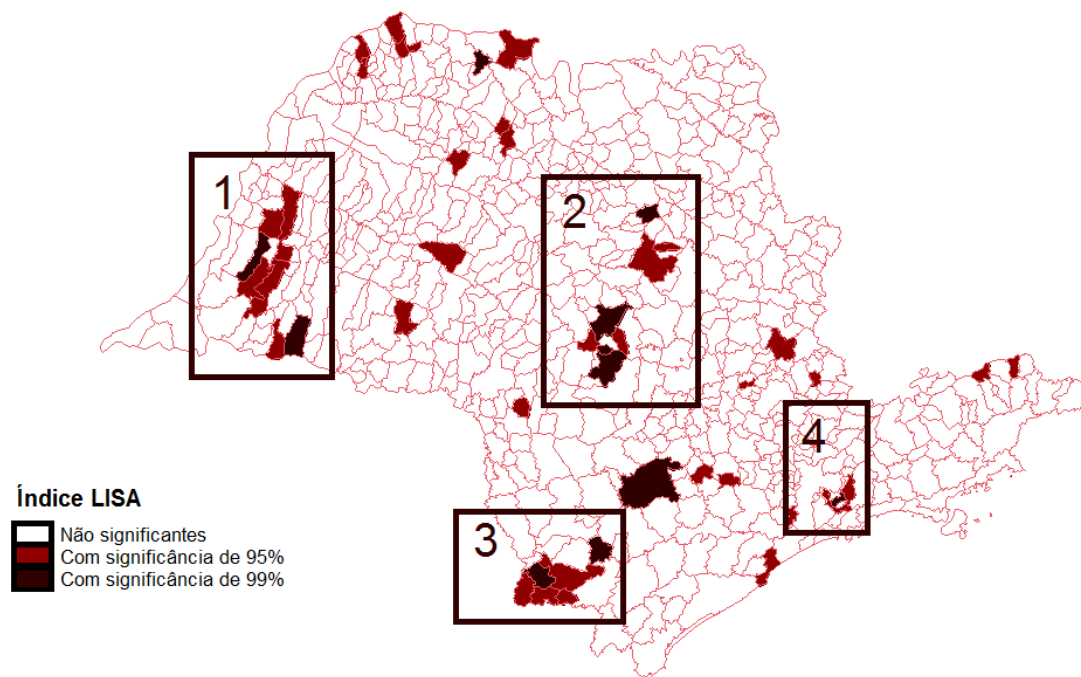
### 3.3 Métodos

Para a elaboração das análises presentes nesse trabalho foram utilizadas análises estatísticas espaciais, com utilização dos Índices de *Moran*.

## 4. Resultados

O resultado do cálculo do Índice Global de *Moran* para a taxa de mortalidade por neoplasia foi de **0,02**, valor próximo de zero, indicando a inexistência de autocorrelação espacial, porém devido ao alto número de áreas estudadas (645 municípios) tornou-se necessário o exame de padrões numa escala de maior detalhe, para verificar se a hipótese de associação espacial do processo verifica-se localmente. Com o cálculo do Índice Local de *Moran*, que encontrou um valor específico de associação espacial para cada município, foi gerado o *LISA map*, que indica as regiões que apresentam correlação local significativamente diferente do restante dos dados. Estas áreas possuem dinâmica espacial própria e merecem análise mais detalhada. Na geração do *LISA map* os índices locais de *Moran* são classificados como não significantes, com significância de 95%, com significância de 99% e 99,9%.

A análise do *LISA map* mostrou diversas áreas de autocorrelação espacial espalhadas pelo estado de São Paulo, contudo podem ser observados alguns agrupamentos de municípios no oeste, centro e sul do estado. Nesse contexto é importante destacar que um agrupamento de municípios indica uma probabilidade grande de se influenciarem de alguma forma no contexto analisado, no caso, o índice de mortalidade por neoplasias. Em análise mais detalhada, principalmente relacionada à quantidade de municípios nos agrupamentos citados, foi possível identificar quatro *clusters* (Figura 2), ou seja, maiores agregados de municípios que apresentam autocorrelação espacial igual ou superior a 95% de significância. Estes *clusters* foram identificados e nomeados de acordo com o município mais significativo da microrregião ou de suas proximidades de acordo com a hierarquia urbana estabelecida pelo Instituto Brasileiro de Geografia e Estatística – IBGE.

**Figura 2** – LISA Map com identificação dos *clusters*

- **Cluster 1 – Microrregião de Presidente Prudente:** No oeste do estado, abrange Junqueirópolis, Dracena, Piquerobi, Emilianópolis, Santo Anastácio, Presidente Bernardes, Narandiba e Taciba.
- **Cluster 2 – Microrregião de Bauru:** Na porção central do estado, abrange Guariba, Santa Lúcia, Araraquara, Ribeirão Bonito, Jaú, Mineiros do Tietê, Macatuba, Igarapu do Tietê, Lençóis Paulista e São Manuel.
- **Cluster 3 – Microrregião de Capão Bonito:** No sul do estado, abrange Guapiara, Apiai, Barra do Chapéu, Itapirapuã Paulista, Ribeira e Itaoca.
- **Cluster 4 – Microrregião de São Paulo:** Próxima à capital do estado, abrange Suzano, Ribeirão Pires, Santo André e São Bernardo.

## 5. Conclusão

O trabalho identificou quatro significativas regiões (*clusters*) de autocorrelação espacial a partir do número de municípios envolvidos e de sua proximidade territorial. De acordo com a metodologia desenvolvida, cada um desses *clusters* apresenta probabilidade acima de 95% de apresentar algum tipo ou fator espacial de correlação. O objetivo da pesquisa era verificar a autocorrelação espacial e identificar os agrupamentos (*clusters*) e, estes apresentam algum padrão ou característica espacial local específica, são regiões com dinâmica espacial própria. Para encontrar os fatores intervenientes das ocorrências de câncer nestes agrupamentos de municípios, deverá haver continuidade na



pesquisa. Isto deve ocorrer com a análise espacial de indicadores, podendo ser geográficos, socioeconômicos, demográficos, entre outros.

## REFERÊNCIAS

- ALMEIDA, C. M. **Aplicação dos Sistemas de Sensoriamento Remoto por Imagens e o Planejamento Urbano e Regional**. Revista Eletrônica de Arquitetura e Urbanismo (USJT). v. 3, p. 98-123, 2010. Disponível em: <[[http://www.usjt.br/arq.urb/numero\\_03/8arq14claudia.pdf](http://www.usjt.br/arq.urb/numero_03/8arq14claudia.pdf)]>. Acesso em: 13 de junho de 2011.
- CÂMARA G., SOUZA R. C. M., FREITAS U. M., GARRIDO J. **SPRING: Integrating remote sensing and GIS by object-oriented data modelling**. Computers & Graphics, 1996.
- CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. V. **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2001. 344 p.
- FERREIRA, A. C. M., RAIJA JUNIOR, A. A. **Análise das taxas de mortalidade por atropelamento dos municípios paulistas utilizando ferramentas de estatística espacial**. Revista dos Transportes Públicos, Ano 32, São Paulo, 2010.
- HOSPITAL ERASTO GAERTNER. **O que é câncer**. Disponível em <<http://www.erastogaertner.com.br/>>. Acesso em: 15 de maio de 2011.
- INCA. **Instituto Nacional do Câncer**. Disponível em <<http://www.inca.gov.br>>. Acesso em: 29 de março de 2011.
- SASSE, A.. **E-Cancer Informações Para Uma Vida Melhor**. Disponível em <<http://andre.sasse.com>>. Acesso em: 20 de setembro de 2011.
- SEADE. **Fundação Sistema Estadual de Análise de Dados**. Governo do Estado de São Paulo. Disponível em: <<http://www.seade.gov.br>>. Acesso em: 25 de junho de 2011.
- SERRANO, R. M.; VALCARCE, E. V. **Técnicas econométricas para El tratamiento de datos espaciales: la econometria espacial**. Barcelona: Ediciones Universidad de Barcelona, 2000.