
**COMPARAÇÃO DOS MÉTODOS SCAN E GETIS&ORD PARA DETECÇÃO DE
CONGLOMERADOS ESPACIAIS DA INCIDÊNCIA DA TUBERCULOSE NO MUNICÍPIO DE JOÃO
PESSOA-PARAÍBA NOS ANOS DE 2004 E 2005**

ANA HERMÍNIA ANDRADE E SILVA
RONEI MARCOS DE MORAES

Universidade Federal da Paraíba - UFPB
Centro de Ciências Exatas e da Natureza - CCEN
Departamento Estatística-DE
Laboratório de Estatística Aplicada ao Processamento de Imagens e Geoprocessamento - LEAPIG
ana_herminia@hotmail.com
ronei@de.ufpb.br

RESUMO – Foi feito um estudo da incidência da tuberculose no município de João Pessoa, Paraíba para os anos de 2004 e 2005. Como detectou-se que a distribuição dos dados é não normal, fez-se uso de uma metodologia não-paramétrica. Com o intuito de escolher o melhor método para a detecção de conglomerados fez-se uso: Varredura Scan e Índice de Correlação de Getis&Ord. O primeiro considera que toda a população deve ter uma incidência homogênea de certo evento, detectando assim a heterogeneidade acima da média para toda a região, já o segundo detecta conglomerados significativos ou não significativos, de valores altos ou baixos. Se pode concluir então que o método de Varredura Scan foi mais eficiente para a detecção de conglomerados espaciais da incidência da tuberculose nos anos de 2004 e 2005 no município de João Pessoa.

ABSTRACT - It was made a study of the tuberculosis incidence in the municipal district of João Pessoa, Paraíba for the years of 2004 and 2005. As it was detected that the distribution of the data is no normal, it was made use of a non-parametrical methodology. With the intention of choosing the best method for the detection of cluster was made use: Spatial Scan and Index of Correlation of Getis&Ord. The first considers that the whole population should have a homogeneous incidence of certain event, detecting like this the heterogeneity above the average for the whole area, already the second detects significant clusters, no significant, of high or low values. It can be concluded then that Spatial Scan method went more efficient for the detection of space conglomerate of the tuberculosis incidence in the years of 2004 and 2005 in the municipal district of João Pessoa.

1 INTRODUÇÃO

A Tuberculose é uma doença infecciosa e contagiosa causada pelo microorganismo *Mycobacterium tuberculosis*, onde sua propagação está ligada às condições de vida da população. Sua proliferação se dá em áreas de grande concentração humana com precários serviços de infra-estrutura urbana onde coexiste fome e miséria. Sua infecção pode ocorrer em qualquer idade, sendo que no Brasil geralmente ocorre na infância. Uma vez infectada, a pessoa pode desenvolver a tuberculose em qualquer fase da vida e todos os órgãos podem ser acometidos pelo bacilo da tuberculose, ocorrendo com mais frequência nos pulmões, gânglios, rins, cérebro e ossos (Ministério da Saúde, 2002).

Estima-se que cerca de 129.000 casos ocorram por ano no Brasil, dos quais cerca de 90.000 são notificados (Ministério da Saúde, 2002), havendo mais de 5 mil mortes anuais (NOGUEIRA et al., 1999). A incidência no país atinge 29,2 pessoas para cada 100.000 habitantes. A região Nordeste é a segunda com maior incidência no

país, 29,9 para cada 100.000 habitantes, juntamente com o Norte (HIJJAR et al., 2001).

Em estudos epidemiológicos é interessante a utilização da análise espacial para detecção geográfica de doenças. Em particular os estudos de detecção de conglomerados são imprescindíveis para a identificação de regiões que apresentam risco elevado em relação à ocorrência de um determinado evento (HIJJAR et al., 2001). Para melhor analisar a distribuição espacial de certa doença pode-se fazer comparações de diferentes métodos de detecção de conglomerados espaciais e optar pelo que melhor traça o perfil da doença.

O objetivo deste trabalho é comparar o método de varredura Scan com o índice de autocorrelação de Geits&Ord para a detecção de conglomerados da incidência da tuberculose no município de João Pessoa nos anos de 2004 e 2005. A utilização dessas técnicas permite que medidas de planos de contenção e direcionamento dos esforços no tratamento desta doença

sejam adotados permitindo melhoria na qualidade de vida da população.

2 METODOLOGIA

Para a análise do comportamento da incidência da tuberculose no município de João Pessoa, com informações fornecidas pela Secretária de Saúde do Município, por meio dos métodos descritos a seguir, foi necessária a projeção da população para os anos de 2004 e 2005. Para tanto foi utilizado o método geométrico (SPIEGELMAN, 1968). Então, calculou-se a taxa de crescimento populacional:

$$r_t = \left[\frac{Pop(t + \Delta t)}{P(t)} \right]^{\frac{1}{\Delta t}} - 1 \quad (1)$$

onde $P(t)$ é a população no tempo t ; $Pop(t + \Delta t)$ é a população no tempo $t + \Delta t$; t é a data base e Δt é o intervalo entre a data base e a data a ser estimada.

A partir da taxa de crescimento obtida projetou-se a população utilizando a seguinte expressão (SPIEGELMAN, 1968):

$$Pop(t + \xi) = P(t)(1 + r)^t \quad (2)$$

onde $Pop(t + \xi)$ corresponde à população projetada para o tempo $t + \xi$.

Para verificar a normalidade da distribuição dos dados de incidência da tuberculose foi aplicado o teste não-paramétrico de Lilliefors (SIELGEL, 1975). Este teste é uma derivação do Teste de Aderência de Kolmogorov-Smirnov que pondera os valores centrais e extremos da distribuição. Tal teste diz respeito ao grau de concordância entre a distribuição de um conjunto de valores amostrais. Com a estatística do teste (p -valor) calculada e com base no nível de significância adotado, pode-se aceitar ou rejeitar a hipótese nula de que os dados seguem uma distribuição normal.

A detecção de conglomerados espaciais tem como objetivo a delimitação de uma região geográfica na qual a hipótese de ocorrência aleatória de um determinado evento pontual é rejeitada (COSTA et al., 2006).

Para a detecção de conglomerados utilizou-se dois métodos com intuito de optar pelo melhor método para a detecção de conglomerados da doença para o período de estudo. Utilizou-se primeiramente o método de Varredura Scan. A estatística Scan é um teste de conglomerados que vem sendo muito utilizado nos últimos anos, devido a sua eficiência em detectar tais regiões (OZONOFF, 2005). O algoritmo do método scan proposto por Kulldorff e Nagarwalla (1995) inicialmente possui um círculo que engloba apenas o centróide de cada sub-região. Em termos gerais, o algoritmo aumenta o raio do círculo englobando um novo centróide até que o círculo reúna todos os centróides da região de estudo (ASSUNÇÃO, 2003). Em outras palavras, tem-se um conjunto de possíveis conglomerados com os raios variando desde a

situação onde somente o centróide em questão esteja na região circular, até um círculo que contenha em seu interior todos os centróides da região suficientes para formação de um conglomerado. Este conjunto de possíveis conglomerados pode ser reduzido se for definido um parâmetro de condição para o raio de busca, de modo que nenhum candidato possua mais que uma dada porcentagem da população (ASSUNÇÃO, 2003).

Sejam os parâmetros (z, p, r), onde z representa os círculos na região de interesse, pertencentes a Z , sendo este o conjunto de todos os círculos, p a probabilidade de um indivíduo qualquer estar dentro de tais círculos seja um caso e r a probabilidade de um indivíduo fora do círculo de interesse seja um caso (COSTA et al, 2005). O método de Varredura Scan é baseado na função de verossimilhança, podendo ser utilizadas as distribuições de Bernoulli ou Poisson. Neste trabalho utilizou-se a distribuição de Poisson. Assim, a hipótese de aleatoriedade completa é dado por:

$H_0 : p = r$ (Cada indivíduo é igualmente provável de se tornar um caso)

$H_1 : p > r$

Os parâmetros da função verossimilhança podem ser estimados por:

$$\hat{p} = \frac{c_z}{n_z} \quad (3)$$

e

$$\hat{r} = \frac{(C - c_z)}{(M - n_z)} \quad (4)$$

onde c_z é o número de casos no círculo z ; n_z é o número de indivíduos em risco no círculo z ; C é o total de casos na região de estudo e M é o total de indivíduos na região de estudo.

A função de verossimilhança para o modelo de Poisson é dado por (COSTA et al, 2005):

$$L(z, p, r) = \frac{\exp(-p * n_z - r * (M - n_z)) * p^{c_z} * r^{(C - c_z)} \prod_i c_i}{C!} \quad (5)$$

onde, c_i é número de casos para $i = 1, 2, \dots, n$.

Para encontrar o conglomerado mais verossímil fixa-se z e calcula-se $p(z)$ e $r(z)$ que maximiza a função de Máxima Verossimilhança. De modo geral, é realizado uma varredura sobre todos os possíveis conglomerados detectados. Logo, o possível aglomerado com Máxima Verossimilhança é a região \hat{Z} definido por (COSTA et al, 2005):

$$L(z, p, r) = \sup_{z, p > r} \frac{\exp(-p * n_z - r * (M - n_z)) * p^{c_z} * r^{(C - c_z)} \prod_i c_i}{C!} \quad (6)$$

Após a definição do aglomerado mais verossímil é associada uma estatística de teste da razão de verossimilhança:

$$k = \frac{L(\hat{z}, p(\hat{z}), r(\hat{z}))}{L_0} \quad (7)$$

onde L_0 é a função de máxima verossimilhança sob a hipótese nula.

A função de verossimilhança para a distribuição de Poisson é dada por (COSTA et al, 2005):

$$L_0 = \frac{C^C (M-C)^{M-C}}{M^M} \quad (8)$$

O outro método utilizado foi o Índice de Autocorrelação de Getis&Ord. Este índice é aplicado a dados que apresentam uma distribuição não-normal, tendo como objetivo fazer uma medida de autocorrelação espacial. Uma restrição para a aplicação deste método, por sê-lo baseado em medidas de distância, é estar destinado apenas as observações positivas (ANSELLIN, 1992).

Os índices de Getis e Ord são estimados a partir de grupos de vizinhos da distância crítica d de cada área i . A distância crítica é formada a partir de uma matriz de proximidade W , onde cada elemento é dito em função da distância crítica, $w_{ij}(d)$.

Getis e Ord propuseram duas funções estatísticas: o índice global $G(d)$ e os índices locais. O índice global é similar às medidas tradicionais de autocorrelação espacial. Os índices locais G_i e G_i^* , que são medidas de associação espacial para cada área i . A partir de um nível de significância, definido como a probabilidade de rejeitar a hipótese nula (existência de autocorrelação espacial), o p -valor é comparado com o índice gerado. Sua avaliação é feita a partir do valor positivo e da significância: o valor positivo e significativo de $G(d)$ indica aglomeração espacial de valores elevados. Em oposição, valores negativos e significantes de $G(d)$ indicam aglomeração espacial de valores pequenos. O índice global de Getis e Ord é dado por:

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) x_i x_j}{\sum_i \sum_j x_i x_j} \quad (9)$$

onde:

X_i é o valor observado na posição i e w_{ij} é um elemento da matriz de proximidade.

Um segundo tipo de estatística sugerido por Getis e Ord é uma medida de associação para cada unidade espacial individual para cada observação i . G_i e G_i^* indicam a extensão a que essa posição é cercada por valores elevados ou por valores baixos para a variável. Este índice relaciona a soma dos valores das posições circunvizinhas à soma dos valores na série de dados ao todo (excluindo a posição considerada). O G_i e G_i^* são dados pelas equações (10) e (11) respectivamente:

$$G_i(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \quad (10)$$

onde a soma é realizada sobre todas as posições excluindo-se a posição i ;

$$G_i^*(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \quad (11)$$

onde a soma é realizada sobre todas as posições, inclusive a posição i .

Por meio do índice local são identificados aglomerados espaciais de valores altos e baixos. Valores padronizados significativos (p -valor menor que 0,05) e positivos indicam uma aglomeração espacial de valores altos. Já valores dos índices padronizados significantes (p -valor menor que 0,05) e negativos apontam para uma aglomeração espacial de valores baixos. As interpretações do índice são efetuadas conforme as indicações abaixo:

1. Negativo*** - Índice local negativo com p -valor menor que 0,005;
2. Negativo** - Índice local negativo com p -valor entre 0,005 e 0,025;
3. Negativo* - Índice local negativo com p -valor entre 0,025 e 0,05;
4. Negativo - Índice local negativa com p -valor acima de 0,05;
5. Positivo - Índice local positivo com p -valor acima de 0,05;
6. Positivo* - Índice local positivo com p -valor entre 0,025 e 0,05;
7. Positivo** - Índices locais com p -valor entre 0,005 e 0,025;
8. Positivo*** - Índice local com p -valor menor que 0,005.

Neste estudo utilizou-se o software R [10] para aplicação das técnicas descritas acima, utilizando para a detecção de conglomerados pelo método Scan e Getis&Ord. Utilizou-se o módulo DCluster, composto por um conjunto de funções para detecção de conglomerados espaciais.

3 RESULTADOS

Para a o estudo da incidência da tuberculose no município de João Pessoa fez-se necessária a projeção da população por bairro para cada ano. Com base na Contagem Populacional de 1996 e no Censo Demográfico de 2000 calculou-se a taxa de crescimento e a população total para o mês correspondente a metade do período anual. Em seguida, calculou-se a proporção populacional para cada bairro, baseando-se nos dados do censo de 2000 e considerando o crescimento populacional constante, obtendo assim a população de cada um dos 66 bairros para o período de estudo.

Para verificar se os dados de incidência de tuberculose apresentavam uma distribuição normal foi aplicado o teste de Lilliefors nos anos de 2004 e 2005. Obteve-se um p-valor menor que $2,2 \times 10^{-16}$, mostrando assim a não normalidade do conjunto. No caso de uma distribuição não normal, a metodologia mais apropriada é a não-paramétrica. Para uma melhor visualização da distribuição espacial da incidência da tuberculose no município de João Pessoa foram gerados mapas coropléticos para os anos do período de estudo. Pode-se observar nas Figuras 1 e 2 que a incidência do ano de 2005 foi maior que a do ano de 2004.

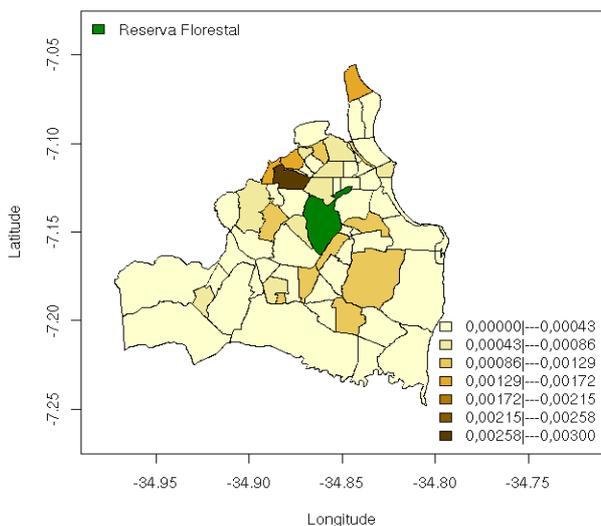


Figura1 - Incidência da tuberculose no ano de 2004.

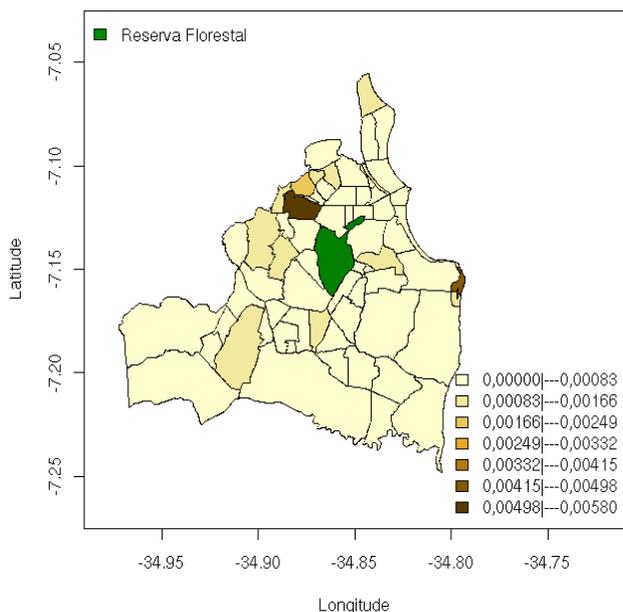


Figura2 - Incidência da tuberculose no ano de 2005.

Com base nos mapas gerados pode-se observar que para ambos os anos do estudo, que as regiões que apresentaram a maior quantidade de bairros com incidência foi a região norte da cidade, sendo que no ano de 2004 a região sul também apresentou muitos bairros com incidência maior que os do restante da cidade. O bairro do Centro, localizado na parte norte da cidade, foi o bairro que apresentou a mais alta incidência para os dois anos do período de estudo, em relação ao restante da cidade.

Para a detecção de conglomerados espaciais fez-se uso de dois métodos, o primeiro deles foi o método de varredura Scan, utilizando a distribuição de Poisson para a detecção de conglomerados espaciais para os anos do estudo. Adotou-se $\alpha = 0,05$ como nível de significância, sendo definido que os candidatos a conglomerados não poderiam superar 10% da população. Foram obtidos os seguintes resultados:

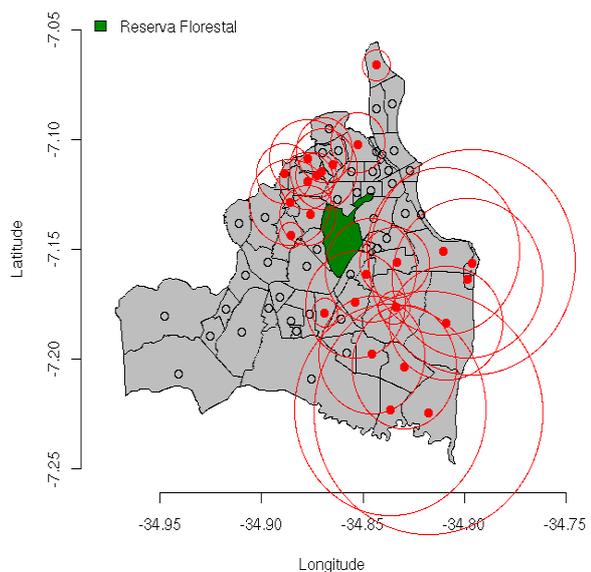


Figura 3 - Método Scan para o ano de 2004.

Como se pode observar nas Figuras 3 e 4, os conglomerados espaciais detectados para ambos os anos coincidem com as regiões com maior número de bairros com incidência, ou seja, a região norte para ambos os anos e região a sul para o ano de 2004. Para a região norte em ambos os anos pode-se observar nas Figuras 3 e 4, baseando-se no tamanho do raio dos círculos formados em cada conglomerado, que a região norte tem conglomerados com a incidência mais alta que o conglomerado da região sul no ano de 2004, o que mostra que a região crítica da incidência da tuberculose é a região norte da cidade e preocupante também na região sul.

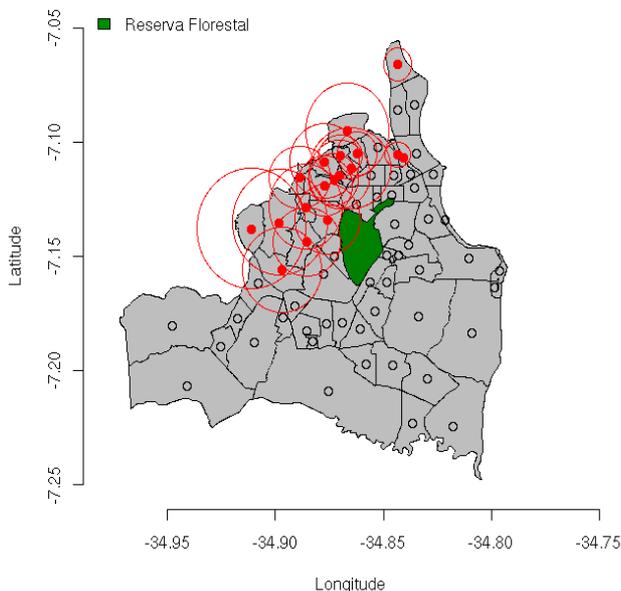


Figura - 4: Método Scan para o ano de 2005.

Com a aplicação do índice de correlação espacial de Getis&Ord foi detectado conglomerado apenas para o ano de 2004. Isto pode ser observado na Figura 5 a seguir:

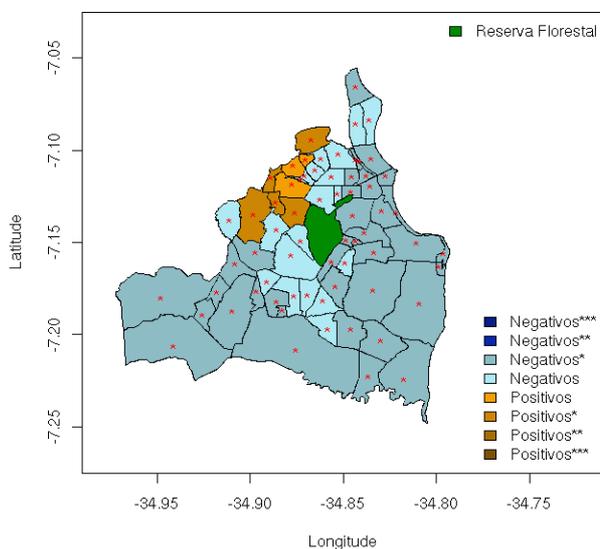


Figura - 5: Método Getis&Ord para o ano de 2004.

Pode-se observar na Figura 4 que com o índice de correlação de Getis&Ord foi detectado um conglomerado significativo de valores altos na região norte do município e um conglomerado de valores baixos nas partes Leste, Sul e Sudeste. Observando a Figura 1 e comparando com a Figura 3 e 5 pode-se observar que houve uma quantidade de bairros na parte sudeste do município que

apresentaram incidência mais alta do que o restante do município, incidência esta detectada apenas pelo método Scan, sendo mais baixa em relação a parte Norte, como mostram os círculos da Figura 3. Já para o ano de 2005 o Método de Getis&Ord não detectou a presença de nenhum conglomerado. Baseando-se na análise dos mapas cadastrais (Figuras 1 e 2) observa-se que a região norte tem alta incidência em relação ao restante do município em ambos os anos, e na região sudeste apenas no ano de 2004, ou seja, estas são as regiões onde é mais provável que existam conglomerados. Essa situação é retratada apenas pela aplicação do método Scan. O método Getis&Ord por sua vez detectou a alta incidência da região norte para o ano de 2004 como conglomerado significativo de valores altos.

4 CONCLUSÕES

Foi feito um estudo sobre a incidência da tuberculose no município de João Pessoa, Paraíba nos anos de 2004 e 2005 com dados fornecidos pela Secretaria de Saúde de João Pessoa. Fez-se necessária a projeção da população para o meio de cada ano de estudo, pelo método Geométrico. Com a aplicação do teste de normalidade de Lilliefors detectou-se que a distribuição dos dados é não normal, sendo então a metodologia mais indicada a ser utilizada a não-paramétrica.

Para uma melhor visualização da distribuição espacial da incidência da tuberculose no município de João Pessoa foram gerados mapas coropléticos para os anos do período de estudo. Com base nos mapas gerados pode-se observar que para ambos os anos do estudo, que as regiões que apresentaram a maior quantidade de bairros com incidência foi a região norte da cidade, sendo que no ano de 2004 a região sul também apresentou muitos bairros com incidência maior que os do restante da cidade.

Aplicou-se dois métodos de detecção de conglomerados, o método de varredura Scan e o índice de correlação espacial de Getis&Ord. Com a aplicação do primeiro detectou-se conglomerados de alta endemicidade na região Norte da cidade para ambos os anos de estudo e de uma endemicidade mais baixa para a região sudeste apenas no ano de 2004. Já o segundo método detectou um conglomerado significativo de valores altos apenas no ano de 2004 na região Norte. Pode-se concluir então que o método de varredura Scan forneceu um resultado mais próximo do esperado pela análise dos mapas cadastrais.

REFERÊNCIAS

Ministério da Saúde, **Manual Técnico para o controle da tuberculose: cadernos de atenção básica/Ministério da Saúde, Secretaria de Políticas de Saúde, Departamento de Atenção Básica, 6ª. revisão ampliada, Brasília, 2002.**

NOGUEIRA, P. A., MALUCELLI, M. I. C., ABRAHAO, R. M. C. M. (2001) "Avaliação das informações de tuberculose (1989 - 1999) de um Centro de Saúde Escola da cidade de São Paulo", **Rev. bras. epidemiol.**, Vol. 4, No 2, pp. 131-138.

HIJJAR, M. A. ,OLIVEIRA, M. J. P. R. ,TEIXEIRA,G,M.(2001)."A tuberculose no Brasil e no mundo". **Boletim de Pneumologia Sanitária**-Vol. 9,No 2.

SPIEGELMAN, M. , "Introduction to demography", Harvard University Press, London, 1968.

SIEGEL, S. "Estatística não-paramétrica para ciências do comportamento", **McGraw-Hill do Brasil**, São Paulo, 1975.

COSTA , Marcelo A. ; SCHRERRER, LUCIANO R. ; ASSUNÇÃO, RENATO M. Detecção de Conglomerados Espaciais com Geometria Arbitrária. IP. **Informação pública**. v. 8, n. 1, 2006.

KULLDORFF, M.; NAGARWALLA, N. Spatial disease clusters: detection and inference. **Statistics in Medicine**, v. 14, p. 799-810, 1995.

ANSELIN, L. "Spatial data analysis with GIS: an introduction to application in the social sciences". Technical Report 92-10. **National Center for Geographic Information and Analysis. University of California**. August, 1992.

OZONOFF, A. ; et al. Cluster detection methods applied to the Upper Cape Cod cancer data. **Environmental Health: A Global Access Science Source**. 4:19. 10.1186/1476-069X-4-19,2005.

ASSUNÇÃO, R. M. Uma Análise de Desempenho dos Métodos Scan e Besag&Newell na Detecção de Conglomerados Espaciais. V **Brazilian Symposium on GeoInformatics**,2003.

COSTA, M. A. ; ASSUNÇÃO, R. M. A Fair Comparison Between the Spatial Scan and the Besag - Newell disease Clustering Tests. **Environmental and Ecological Statistics**, Vol. 12, 301-319, 2005.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Disponível em: <http://www.R-project.org/>. Acesso em: 5 dezembro de 2007.