

Estatística Espacial como Ferramenta do Planejamento Urbano

Análise de Áreas: Setores Censitários, Zonas OD, Distritos e Subprefeituras.

Carlos Paiva

Este resumo foi desenvolvido com base nos seguintes estudos:

Dissertação de Mestrado - Frederico Roman Ramos - <http://www.dpi.inpe.br/teses/fred/>

Orientadores: Gilberto Câmara e Antônio Miguel Vieira Monteiro

Análise Espacial de Áreas - <http://www.dpi.inpe.br/gilberto>

Gilberto Câmara, Marília Sá Carvalho, Oswaldo Gonçalves Cruz, Virgínia Corrêa

A *dependência espacial* é uma característica inerente à representação de dados através de subdivisões territoriais. Ela pode ser entendida como a tendência a que o valor de uma variável associada a uma determinada localização assemelhe-se mais ao valor de suas amostras vizinhas do que ao restante das localizações do conjunto amostral. O conceito de vizinhança adotado pode ser formulado com base na propriedade de proximidade, (zonas fronteiriças, distancia linear, etc.), como também na conectividade (tempo de viagem, número de viagens, etc.).

A necessidade de quantificação da dependência espacial presente num conjunto de geodados levou ao desenvolvimento da chamada *estatística espacial* cuja característica principal é seu foco em inquirir *padrões espaciais* de lugares e valores, identificando: a *associação espacial existente* entre eles e a variação sistemática do fenômeno por localização Anselin (1992). Há também que se destacar, que as associações espaciais identificadas como significativas, através desta metodologia, estão necessariamente associadas à resolução espacial dos dados de entrada, e que uma resolução diferenciada pode revelar outros padrões espaciais.

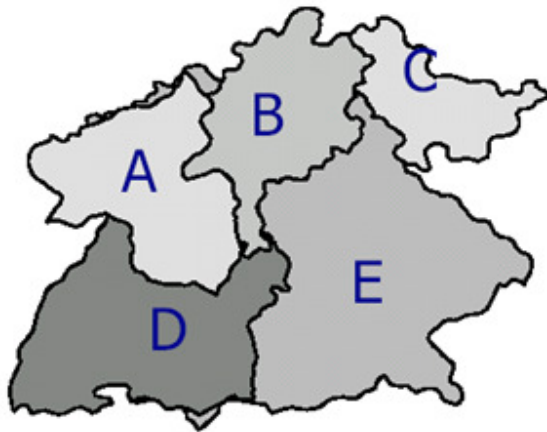
Estatística Espacial

Média Móvel Espacial

Uma forma simples de explorar a variação da tendência espacial dos dados é calcular a média dos valores dos vizinhos (*Média Móvel Espacial*). Isto reduz a variabilidade espacial, pois a operação tende a produzir uma superfície com menor flutuação que os dados originais. A média móvel $\hat{\mu}_i$ associada ao atributo z_i , relativo à i -ésima área, pode ser calculada a partir dos elementos w_{ij} da matriz normalizada de proximidade espacial W , tomando-se simplesmente a média dos vizinhos:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij} z_j$$

A matriz de vizinhança comumente utilizada para geodados zonais define vizinhança a partir da propriedade topológica de contiguidade, assim W é uma matriz binária (0,1), onde 1 está associado às zonas com fronteiras em comum e 0 àquelas sem esta propriedade. Como a matriz de vizinhança é utilizada em cálculos de indicadores de análise exploratória, por conveniência, ela é muitas vezes utilizada normalizada por linha, ou seja, com a soma dos ponderadores de cada linha igual a um (Câmara et al.,2002).



	A	B	C	D	E
A	0	0,5	0	0,5	0
B	0,25	0	0,25	0,25	0,25
C	0	0,5	0	0	0,5
D	0,33	0,33	0	0	0,33
E	0	0,33	0,33	0,33	0

Exemplo de divisão zonal com matriz de vizinhança normalizada por linha associada à propriedade de contigüidade.

O índice de Moran (I)

A dependência espacial pode ser medida de diferentes formas. O índice de Moran (I) é a estatística mais difundida e mede a autocorrelação espacial a partir do produto dos desvios em relação à média. Este índice é uma medida global da autocorrelação espacial, pois indica o grau de associação espacial presente no conjunto de dados.

De uma forma geral, o índice de Moran presta-se a um teste cuja hipótese nula é de independência espacial; neste caso, seu valor seria zero. Valores positivos (entre 0 e +1) indicam para correlação direta e negativos, (entre 0 e -1) correlação inversa. Uma vez calculado, é importante estabelecer sua validade estatística. O mais comum é um teste de pseudo-significância. Neste caso, são geradas diferentes permutações dos valores de atributos associados às regiões; cada permutação produz um novo arranjo espacial, onde os valores estão redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, pode-se construir uma distribuição empírica de I . Se o valor do índice I medido originalmente corresponder a um “extremo” da distribuição simulada, então se trata de valor com significância estatística.

$$I = \frac{n}{W} \left(\frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \right) \text{ para } i \neq j$$

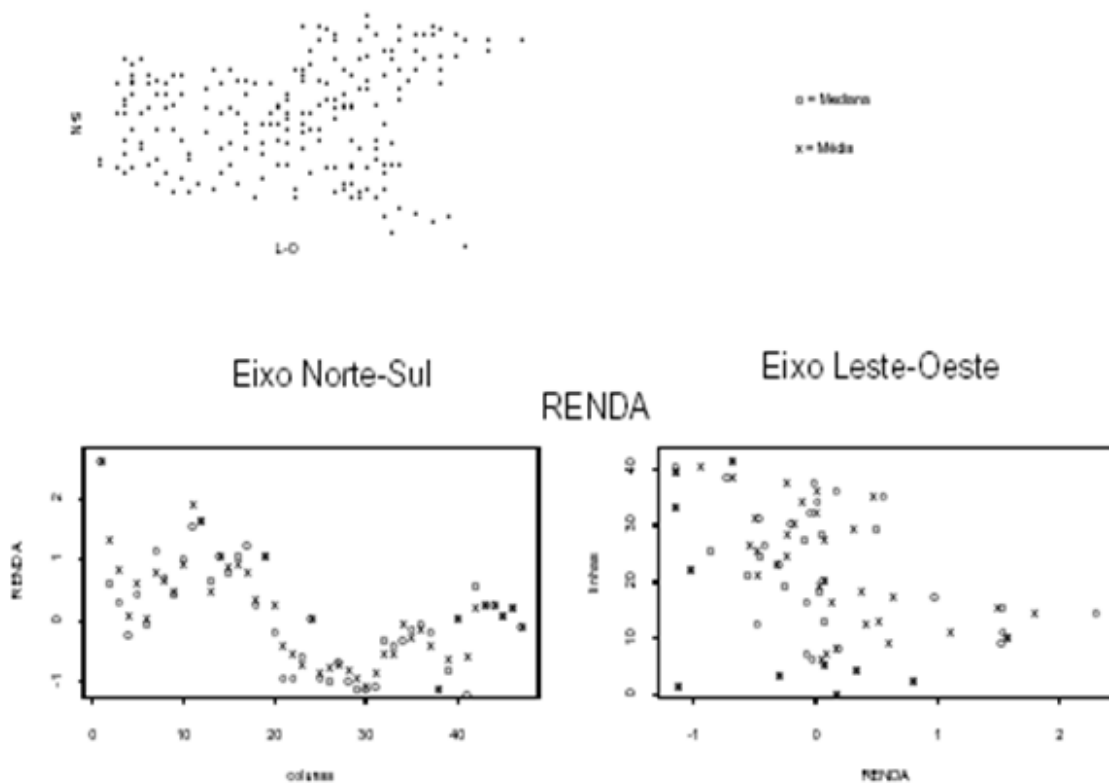
onde:
 n é o número de observações
 w_{ij} é o elemento na matriz de vizinhança para o par i e j
 W é a soma dos ponderadores da matriz
 z_i e z_j são desvios em relação à média ($z_i - z$), ($z_j - z$)
 z é a média

Neste caso, a correlação será computada apenas para os vizinhos de primeira ordem no espaço, conforme estabelecido pelos pesos w_{ij} . O mesmo cálculo feito para matrizes de proximidade de maior ordem permite estimar a função de autocorrelação para cada ordem de vizinhança (ou “lag”).

$$I^{(k)} = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(k)} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^N (z_i - \bar{z})^2}$$

A hipótese implícita do cálculo do índice de Moran é a estacionariedade de primeira e segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. Os gráficos de médias e medianas segundo linhas e colunas permitem explorar simultaneamente a presença de tendência (não-estacionariedade de primeira ordem), e não-estacionariedade de segunda ordem, onde a variância e a covariância entre vizinhos não se mantém constante.

Para construir estes gráficos, utiliza-se as coordenadas dos centróides das áreas, aproximando-as para um espaçamento regular de forma a montar uma matriz. Calcula-se então as médias e as medianas do indicador ao longo das linhas (eixo Leste-Oeste) e colunas (eixo Norte-Sul) desta matriz. Esta técnica permite identificar a flutuação das medidas ao longo de duas direções, sugerindo a presença de valores discrepantes quando a diferença entre estas é grande (não-estacionariedade de segunda), e a tendência (não-estacionariedade de primeira ordem), ao longo de uma direção quando os valores variam suavemente.



Médias (x) e Medianas (o) para Renda Média do Chefe de Família na Ilha do Governador - RJ

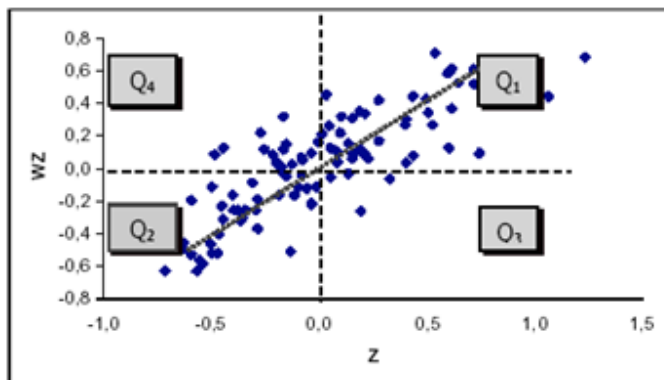
No eixo Norte-Sul (colunas) pode-se observar que a renda média do chefe da família apresenta tendência variável, bem menor no centro da região. No eixo Leste-Oeste (linhas), também parece haver algum deslocamento para valores mais altos no sentido leste, mas o descolamento de médias (x) e medianas (o) sugere a presença de valores extremos dos indicadores. A variação na média dos indicadores na região está, aparentemente, dividida entre as duas direções analisadas.

Algumas variações deste modelo são o teste *C* de Geary e o teste *Ipop*. O indicador *C* de Geary assemelha-se ao variograma, e o *I* de Moran ao correlograma. Deixaremos estes indicadores para estudo em próximo artigo.

Diagrama de Espalhamento de Moran

O diagrama de espalhamento de Moran (*Moran Scatterplot Map*) é uma maneira adicional de visualizar a dependência espacial. Construído com base nos valores normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão), permite analisar o comportamento da variabilidade espacial.

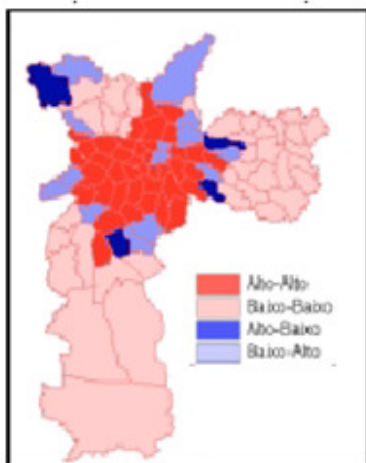
A idéia é comparar os valores normalizados do atributo numa área com a média dos seus vizinhos, construindo um gráfico bidimensional de z (valores normalizados) por wz (média dos vizinhos, também normalizadas), que é dividido em quatro quadrantes, como mostrado abaixo.



Os quadrantes podem ser interpretados como:

- Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.
- Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos, indicando pontos de transições entre diferentes padrões espaciais ou pontos de não estacionariedade do atributo.

O índice de Moran I é equivalente ao coeficiente de regressão linear que indica a inclinação da reta de regressão de wz em z . O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa temático bidimensional, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento.



Mapa de Espalhamento de Moran para o índice de exclusão/inclusão social da cidade de São Paulo, censo 1991.

Nesta figura, “Alto-Alto”, “Baixo-Baixo”, “Alto-Baixo” e “Baixo-Alto” indicam, respectivamente, os quadrantes Q1, Q2, Q3 e Q4, mostrados na Figura acima. Nota-se uma forte polarização centro-periferia e observa-se que os distritos localizados nos quadrantes Q3 e Q4 (indicados pela cor azul) podem ser entendidos como regiões de transição entre o centro da cidade (que tende a apresentar valores positivos do índice de exclusão/inclusão social) e as duas grandes periferias de São Paulo (zona Sul e zona Leste).

Estatística Espacial Local

A *estatística espacial local* foi desenvolvida para quantificar o grau de associação espacial a que cada localização do conjunto amostral está submetida em função de um modelo de vizinhança preestabelecido. Convencionou-se chamar de *Indicadores Locais de Associação Espacial* (LISA) os operadores estatísticos com esta finalidade. Anselin (1995) aponta que existe uma proporcionalidade direta entre o valor da autocorrelação global e os valores das autocorrelações locais, ele demonstra que os LISAs permitem a decomposição dos indicadores globais em contribuições individuais, indicando porções territoriais de não *estacionariedade* e identificando aglomerados (*clusters*) significativos de valores semelhantes em torno de determinadas localizações. Dentre os LISAs mais difundidos estão o Índice Local de Moran (I_i) e as Estatísticas G_i e G_i^* .

No índice local de Moran a autocorrelação espacial é calculada a partir do produto dos desvios em relação à média como uma medida de covariância, dessa forma, valores significativamente altos indicam altas probabilidades de que haja locais de associação espacial tanto de regiões com altos valores associados como com baixos valores associados.

$$I_i(d) = \frac{(x_i - \bar{x})}{s^2} \sum_j w_{ij}(d)(x_j - \bar{x}) \quad \text{para } j \neq i$$

Nas estatísticas G_i e G_i^* diferentemente, o cálculo é realizado pela soma simples das amostras vizinhas em relação a uma determinada posição i , sendo que no caso do índice G_i , o valor da localização em estudo i não é incluído na somatória e, no caso do índice G_i^* , sim. Portanto nestas, valores significativamente altos indicam locais de associação espacial de amostras de alto valor e valores significativamente baixos indicam agrupamentos de amostras de baixos valores.

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad \text{para } j \neq i$$

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad \text{para todos os } j$$

onde para os dois conjuntos de formulas acima temos:

$w_{ij}(d)$ é o ponderador na matriz de vizinhança W para o par i e j
 d é a medida de distância estabelecida pelo modelo de vizinhança
 x_i e x_j são valores encontrados na posição i e suas vizinhas js
 \bar{x} é a média amostral global

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{Frota, 2002}).$$

s^2 é a variância amostral global. (média aritmética dos desvios quadráticos)

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{34} (x_k - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_{34} - \bar{x})^2}{n-1} \quad (\text{Frota, 2002}).$$

A significância referida em ambos os casos dependerá da suposição de alguma distribuição estatística ou da construção de uma pseudo-distribuição empírica através de permutação (Getis e Ord,1992). A técnica de permutação permite estabelecer significância estatística empírica do resultado obtido. Esta significância empírica é obtida a partir de uma distribuição gerada pela repetida substituição aleatória dos n valores pertencentes ao conjunto amostral nas i -ésimas localizações e calculando-se novos resultados para cada novo arranjo. Assim, tem-se a possibilidade de comparação do resultado obtido sob arranjo espacial original e os n resultados obtidos nos arranjos gerados aleatoriamente (Anselin,1992) (Câmara et al.,2002).

As estatísticas espaciais locais se colocam como ferramentas importantes por sua capacidade de inferir padrões locais e globais da distribuição espacial das variáveis georreferenciadas.

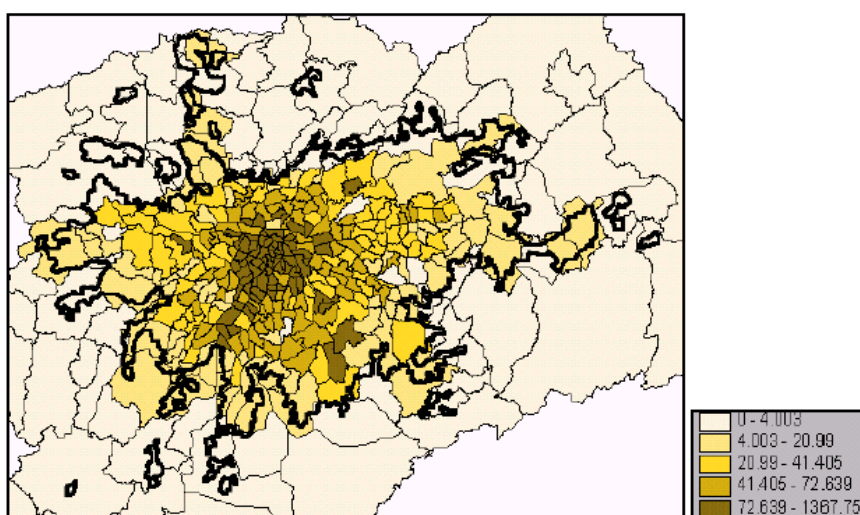
Uma aplicação direta dos LISAs é inferir sobre as *áreas de influência* de determinadas variáveis, indicando a extensão, dentro de um conjunto de localizações, de padrões espaciais locais onde há uma tendência em se encontrar valores similares para essas amostras. A capacidade de detecção de agrupamentos significativos de amostras de valores próximos em torno de cada localização permite identificar os pontos onde tais características predominam, apontando assim, potenciais territórios desenhados a partir destas características.

Uma outra aplicação está relacionada à estrutura intra-urbana que é a noção de topologia intra-urbana. Associadas a este conceito estão às características distributivas identificadas como *persistências e transições*, ou seja, expressões do desenho distributivo global da variável no espaço

que revelam, além das concentrações, locais destacados a partir das diferenças, informação muitas vezes relevante na análise. Estes pontos de autocorrelação espacial negativa podem ser interpretados como polígonos estranhos a sua vizinhança. A possibilidade de identificação destes padrões de persistência e transição das características associadas às variáveis georreferenciadas traz uma expressão possível da topologia intra-urbana, relacionada às localizações relativas dos elementos componentes da estrutura e suas articulações espaço-funcionais.

Exemplificando - Pesquisa Origem Destino 1997 (OD97)

A variável selecionada foi o total de empregos por zona, dividido por sua área total, ou seja, a densidade de empregos (emp_área). Esperava-se com esta variável detectar agrupamentos de zonas com altas densidades de empregos que pudessem apontar para o centro e subcentros da estrutura intra-urbana de São Paulo. A figura abaixo apresenta as zonas OD97 intra-urbanas classificadas por quantís.



Densidade de empregos classificados por quantís;

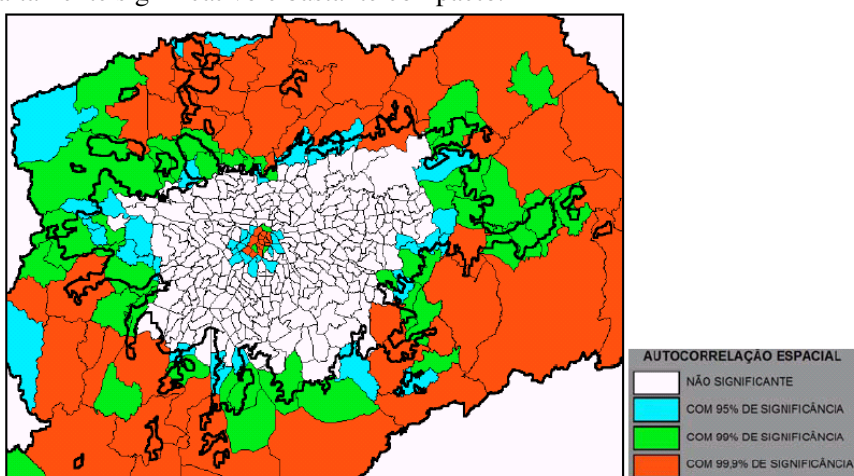
Neste estudo da densidade de empregos na região metropolitana três ferramentas exploratórias foram utilizadas:

- Mapa de Significância do Índice Local de Moran,
- Mapa das Estatística G_i^* Classificadas por Desvios Padrões,
- e Mapa de Espalhamento de Moran.

No experimento, a matriz de vizinhança foi definida pela propriedade topológica de contiguidade e utilizada normalizada nos cálculos processados. Todas as visualizações são apresentadas sobrepostas ao contorno da mancha urbana extraído da imagem landsat-TM gerada em 1997, assim direciona-se a interpretação aos polígonos urbanizados, mais densamente ocupados. Esta forma de apresentação deve-se ao fato de que os polígonos mais externos da RMSP são pouco representativos da estrutura intra-urbana, eles podem ser afetados por instabilidades decorrentes da baixa densidade demográfica como é sugerido em Análise de Dados Agregados em Áreas.

A decisão de não aplicar filtros de estimação bayesiana antes do cálculo dos LISAs foi em função do reconhecimento da influência que estes estimadores, baseados em médias globais ou locais conhecidas a priori, exercem nas medidas de autocorrelação espacial sobretudo naqueles polígonos com baixa representatividade populacional. Assim, optou-se por preservar as autocorrelações espaciais originais observadas para o cálculo dos LISAs e direcionar a análise para os polígonos mais representativos da estrutura intra-urbana, ou seja, os internos à mancha.

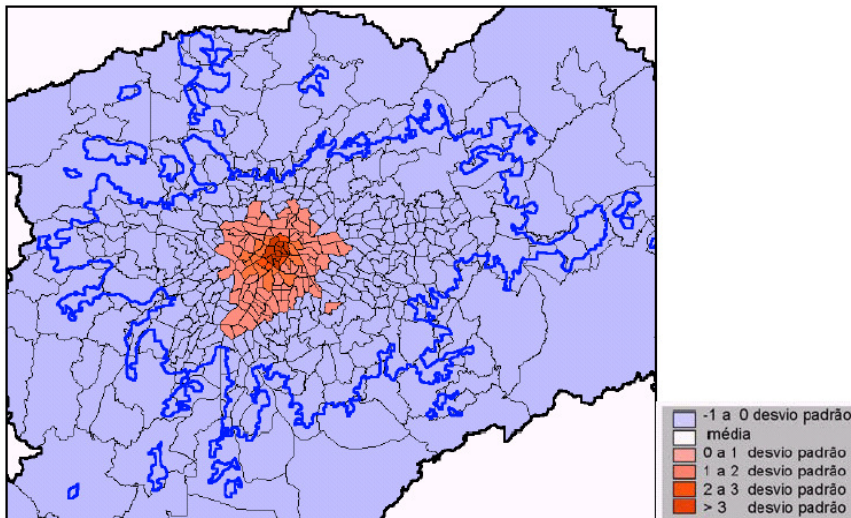
As análises preliminares realizadas através do índice de Moran global sobre a variável emp_area apontaram com significância de 99,9% após 999 permutações a existência de um padrão de autocorrelação espacial. De fato, as análises de autocorrelação espacial local efetuadas revelam um padrão de extrema concentração da densidade de emprego nas zonas mais centrais da cidade. O mapa de significância abaixo do índice local de Moran calculado indica que este agrupamento é altamente significativo e bastante compacto.



Mapa de significância dos resultados obtidos para índice local de Moran após 999 permutações.

A área mais central da cidade, com maior significância, 99,9%, estende-se da região da Luz, ao norte, até a região da Avenida Paulista, a sudoeste. Com significância menor, 95%, um anel de polígonos circunda este núcleo e abrange no eixo leste-oeste desde a zona leste próxima, Brás e Bresser, até Pinheiros; e no eixo noroeste-sudeste, de Perdizes a Vila Mariana. Afora este agrupamento, quase todos os outros polígonos internos a mancha urbana não apresentam significativa autocorrelação espacial, indicando não haver um padrão específico da densidade de emprego nestas áreas. Nos polígonos mais externos, a significativa autocorrelação foi atribuída devido à persistência do padrão de baixas densidades de emprego relacionadas às áreas rurais externas a estrutura.

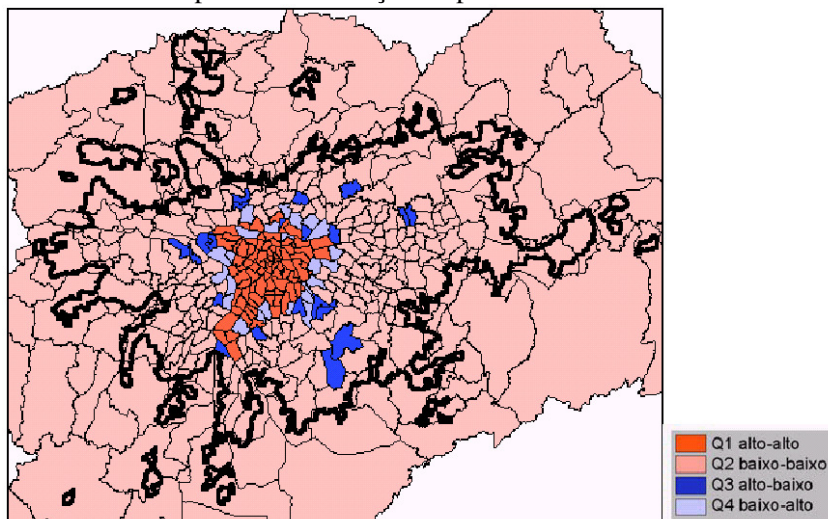
O mapa dos índices G_i^* normalizados é uma ferramenta interessante pela simplicidade de construção e interpretação, é uma maneira direta de se avaliar o comportamento global das variáveis sobre o espaço. No mapa abaixo, a classificação por desvios também aponta para os polígonos centrais como positivamente autocorrelacionados e apresenta um terceiro anel mais externo positivamente correlacionados que pode ser associado à extensão do centro expandido definido pelo anel viário (Marginais - Bandeirantes - Tancredo Neves - Salim F. Maluf), sendo que na direção da Zona Norte e de Santo Amaro este limite é ultrapassado.



Os índices G_i^* normalizados classificados por desvios em relação à média.

O cálculo do G_i^* é uma soma simples de vizinhos normalizada, assim, sua visualização classificada por desvios em relação à média permite uma interpretação direta da intensidade e do sinal da associação espacial. Na variável normalizada, a média tende a zero, assim a representação se faz por números de desvios padrões em relação ao zero em valor positivo, indicando os locais onde esta soma foi maior que zero e negativos em casos contrários.

O mapa de espalhamento de Moran (Figura abaixo) permite a visualização dos polígonos relacionados aos pontos de transição de padrão.



Mapa de espalhamento de Moran obtido para a variável densidade de empregos.

Estes polígonos indicam as áreas onde diferentes padrões se encontram, imprimindo uma instabilidade na correlação espacial local identificada no gráfico de espalhamento. Chamam a atenção os polígonos identificados como Q3, valor alto inserido em vizinhança com média baixa, fora do núcleo central compacto. Uma análise mais atenta revelou que estes polígonos estão associados a importantes subcentros na estrutura territorial da metrópole, caso do centro de Guarulhos, São Caetano, Santo André, São Bernardo do Campo e Osasco, além das zonas OD97

correspondentes a São Miguel Paulista e a cidade Nitro-Operária também indicada como pólos de empregos inseridos na vizinhança de baixa densidade de empregos na zona leste.

Em conjunto, estas três ferramentas exploratórias apontaram evidências de que os territórios do trabalho, componentes essenciais da estrutura intra-urbana, se configuram com um núcleo compacto com alta densidade de empregos nas zonas centrais, circundado por anéis de densidade decrescente desta atividade. Este núcleo destaca-se do padrão predominante das outras zonas OD97 internas a mancha urbana, apontados como áreas de baixas densidades pelo mapa Gi*.

Dentro destes territórios com poucos empregos algumas instabilidades pontuais foram detectadas pelo mapa de espalhamento de Moran, revelando-se como os subcentros da estrutura intra-urbana. Desenha-se assim, a expressão matemática da configuração territorial das atividades empregatícias intra-urbanas da cidade de São Paulo. A extrema concentração dos empregos na cidade é reconhecidamente uma das principais causas dos grandes problemas de acessibilidade que a cidade vem enfrentando, pois ao gerar enormes deslocamentos diários para uma grande faixa da população, impõe perdas na sua produtividade e qualidade de vida.

É importante observar que existem softwares na internet que permitem a usuários não familiarizados a cálculos estatísticos, a utilização destas metodologias de análise espacial de modo bastante simplificado. Software GeoData - <http://www.csiss.org/clearinghouse/GeoDa/> e Terra View - <http://www.dpi.inpe.br/terraview/index.php>.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anselin, L. Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27, n.2, 93/115, 1995.
- Anselin, L. SpaceStat tutorial [on line], 1992. <<http://www.spacestat.com>>, 1999
- Câmara, G., Carvalho, M.S., Cruz, O.G., Correa, V. *Análise Espacial de Áreas*, 2002, (*mimeo*)
- Frota, A. *Aprenda Estatística Através da Pesquisa*. [on line], 2002. <<http://usuarios.lycos.es/EQP/index.htm>>.
- Ramos, R. F. *Análise Espacial de Estruturas Intra-Urbanas: O caso de São Paulo*, dissertação de Mestrado – INPE, 2002