



## PROCEDIMENTO METODOLÓGICO PARA PROPOSTA DE INDICADOR DE ASSOCIAÇÃO ESPACIAL GLOBAL ATRAVÉS DE CONCEITOS VARIOGRÁFICOS

Cláudia C. B. R. Naizer

Cira S. Pitombo

Universidade de São Paulo

Escola de Engenharia de São Carlos/Departamento de Engenharia de Transportes

### RESUMO

Entre as ferramentas de análise espacial exploratória de dados, encontram-se indicadores globais de associação espacial e o cálculo e modelagem de semivariogramas. Neste trabalho pretende-se utilizar conceitos do semivariograma para proposta de indicador global. Utilizaram-se os valores de escolha modal de dois bancos de dados: uma cidade fictícia e o bairro Santa Felícia, cidade de São Carlos – SP. Os dados de probabilidade de escolha modal foram analisados a distâncias fixas, denominadas vizinhanças, para as quais foram calculados e ajustados semivariogramas experimentais e teóricos para as direções principal e ortogonal. Os valores pontuais obtidos em ambos os semivariogramas são os indicadores globais para as diversas vizinhanças. O indicador foi associado a um teste de pseudo-significância, de forma a comprovar associação espacial dos dados. Constatou-se que na maioria dos casos em que a hipótese de dependência espacial não foi refutada pelo teste aqui proposto, foram obtidos valores altos do I de Moran.

### ABSTRACT

Among the spatial data analysis tools, there are global indicators of spatial association and variographic analysis. The aim of this paper is to use the concept of semivariogram for a global indicator proposal. The travel mode choice values of two databases were used: a hypothetical city and the region of Santa Felícia, city of São Carlos - SP. The mode choice probability data were analyzed at fixed distances, called neighborhoods, for which the experimental and theoretical semivariograms were calculated and adjusted for the principal and orthogonal directions. The point values obtained in both semivariograms are global indicators for the various neighborhoods. The indicator was associated with a pseudo-significance test, in order to prove the spatial dependence of the data. It was found that in most of the cases in which the hypothesis of spatial dependence was not refuted by the proposed test, high values of Moran's I index were obtained.

### 1. INTRODUÇÃO

A demanda por transportes é espacialmente dependente. Os indivíduos decidem as origens de suas viagens e seus destinos quando escolhem onde morar, trabalhar, estudar e realizar lazer, bem como escolhem a rota entre essas localidades. Viagens ocorrem no espaço, assim, estão intimamente ligadas a atributos geográficos (Loureiro *et al.*, 2006). Abordagens atuais enfatizam a necessidade de inclusão de componentes espaciais, pois sabe-se que diversos componentes do processo de modelagem de transportes são espacialmente dependentes, o que contraria as premissas de independência dos dados dos métodos tradicionais (Sener *et al.*, 2010).

Diversos autores desenvolveram indicadores de análise espacial, de forma a comprovar a existência de autocorrelação espacial e quantificar relações de dependência espacial. A maioria dos índices diz respeito à similaridade entre valores da mesma variável em diferentes localizações e é aplicável predominantemente a variáveis quantitativas, como é o caso do índice I de Moran global (1950).

Dentre as técnicas de análise espacial também se encontra a Geoestatística, que parte do pressuposto de mensurar a dissimilaridade dos valores da mesma variável em diferentes pontos de localizações diferentes, através do semivariograma. Enquanto os indicadores espaciais, comumente conhecidos na literatura, apresentam o resultado na forma de um número, a geoestatística reflete o fenômeno de autocorrelação espacial através do gráfico semivariograma, que relaciona valores de semivariância e distância entre pontos. A técnica é aplicável a variáveis



quantitativas e qualitativas, desde que transformadas em dados binários.

Tanto os índices de análise espacial quanto o semivariograma consideram distâncias fixas relevantes para os cálculos, excluindo os demais pontos que não pertencem a tais distâncias. Porém, no caso dos indicadores, apenas a primeira vizinhança (distância previamente especificada) é relevante, sendo possível perder eventualmente resultados significativos de vizinhos a distâncias pouco maiores. A análise de todas as vizinhanças possíveis até a estacionariedade espacial é possível através do semivariograma.

O presente projeto utiliza os valores obtidos através de semivariogramas experimentais e teóricos como indicadores de associação espacial global, obtendo um único valor que caracterize e quantifique a autocorrelação espacial. A utilização da Geoestatística possibilita trabalhar com diferentes tipos de dados, ao contrário dos indicadores tradicionais. Além disso, a ferramenta oferece subsídios para melhor investigação de direção angular (e tolerância) para análise da maior variabilidade espacial de uma variável regionalizada.

Desta forma, a contribuição do presente trabalho está tanto na melhor investigação da estrutura espacial das variáveis de demanda por transportes (direção principal e ortogonal, vizinhanças e estacionariedade), na possibilidade de uso de diferentes tipos de variáveis, bem como na proposta de um teste estatístico de hipótese, associado ao indicador, para teste de dependência espacial em diversas vizinhanças possíveis.

## 2. ANÁLISE ESPACIAL

Tobler (1970) estabeleceu a “Primeira Lei da Geografia”, a qual descreve o conceito de dependência espacial, abordando a tendência que o valor de uma variável no espaço tem de ser mais parecido com o valor de amostras vizinhas do que de amostras distantes.

Moran (1950) propôs que as variáveis fossem padronizadas através da subtração da média simples e então multiplicadas por um fator. Cliff e Ord (1972) generalizaram a estatística através da derivação do teste para correlação espacial em um modelo de regressão linear.

Poucos anos depois, Geary (1954) desenvolveu a estatística  $c$ , baseada na dissimilaridade espacial. Enquanto a estatística de Moran (1950) se assemelha a um correlograma (gráfico de correlação por distância entre pontos) a estatística  $c$  de Geary é similar a um semivariograma (gráfico de semivariância por distância entre pontos). Dessa forma, as estatísticas são inversamente relacionadas. A estatística de Moran é uma medida de autocorrelação global, enquanto a estatística  $c$  de Geary é mais sensível a autocorrelação local. A estatística de Moran possui um teste de hipótese mais potente do que a estatística  $c$  de Geary (Cliff e Ord, 1981).

Getis e Ord (1992) desenvolveram a família de estatísticas G, composta pela estatística global  $G(d)$  e pelas estatísticas locais  $Gi(d)$  e  $Gi^*(d)$ . No estudo, Getis e Ord (1992) recomendam que a técnica seja utilizada em conjunto com o índice I de Moran.

A princípio, as técnicas de análise espacial de índices seguiram paralelamente ao desenvolvimento da geoestatística. Ambos são utilizados para exploração e inferência de dados espacialmente correlacionados, porém os resultados dos métodos são diferentes. Os índices retornam um único número e estão atrelados a um teste de significância, de forma a descartar a hipótese de aleatoriedade espacial dos dados. Na geoestatística, as semivariâncias são dadas segundo a distância entre pontos na forma do gráfico semivariograma, não sendo possível realizar um teste de significância sobre o diagrama.

Fortin *et al.* (1989) investigaram a autocorrelação espacial aplicada à ecologia, utilizando coeficientes de autocorrelação espacial e geoestatística. Concluiu-se que as técnicas podem ser



aplicadas de forma complementar, para compreensão do fenômeno estudado através de diferentes perspectivas.

## 2.1. Indicadores Globais de Autocorrelação Espacial

Indicadores globais de autocorrelação espacial retornam um único valor como medida de associação espacial do conjunto de dados analisado. Os principais índices globais são apresentados a seguir na Tabela 1.

**Tabela 1:** Indicadores Globais de Autocorrelação Espacial. Em que:  $w_{ij}(d)$  são os elementos da matriz de pesos para a ordem de vizinhança ou lag d;  $x_i$  região em análise,  $x_j$  elementos considerados como vizinhança e utilizados para o cálculo;  $\bar{x}$  média da variável; n número de elementos.

Estatística	Equação
Índice I de Moran	$I(d) = \frac{n \sum_{i=1}^n \sum_{j=0}^n w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ (1)
Estatística c de Geary	$c = \frac{n-1}{2 \sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)} \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)(x_i - x_j)^2}{\sum_{i=0}^n (x_i - x_j)^2}$ (2)
Estatística G(d)	$G(d) = \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)x_i x_j}{\sum_{i=0}^n \sum_{j=0}^n x_i x_j}$ (3)

Para a aplicação dos indicadores, deve-se definir a matriz de proximidade espacial, a qual estima a variabilidade espacial dos dados, podendo ser determinada por contiguidade (áreas), distância (pontos) ou k-vizinhos mais próximos (pontos). A matriz de proximidade espacial W é quadrada, e composta pela relação entre cada par de áreas ou pontos i e j.

O índice de Moran varia de -1 a +1, valores negativos indicam autocorrelação espacial negativa, valores próximos de zero indicam ausência de autocorrelação espacial e valores positivos indicam autocorrelação espacial positiva (Moran, 1950). A estatística c de Geary (1954) varia de 0 a 2, sendo que 0 a 1 corresponde a autocorrelação espacial positiva e 1 a 2 autocorrelação espacial negativa. A estatística global G(d), representa um índice de concentração espacial, pois identifica apenas autocorrelação espacial positiva, classifica os vizinhos em “hot spots” (variável com valor alto vizinhos com valor alto) e “cool spots” (variável com valor baixo e vizinhos com valor baixo) (Getis e Ord, 1992).

## 2.2. Geoestatística

A Geoestatística analisa conjuntos de dados espaciais através de uma interpretação estocástica. Diferentemente da Estatística Clássica, as observações são interpretadas em conjunto de forma a caracterizar o padrão espacial do fenômeno fielmente (Druck *et al.*, 2004). Para compreensão do indicador proposto neste estudo, são conceituados em seguida termos importantes relativos à ferramenta, tais como Variáveis Regionalizadas, Semivariogramas (experimentais e teóricos).

Matheron (1971) desenvolveu a teoria das variáveis regionalizadas, as quais são funções numéricas com dependência espacial e que não podem ser explicadas através de funções determinísticas (Druck *et al.*, 2004). Atualmente, é possível aplicar técnicas geoestatísticas a variáveis quantitativas (contínuas e discretas) e qualitativas (nominais, ordinais e binárias), com a ressalva que os dados qualitativos nominais devem ser transformados em variáveis binárias (0-1) para que o procedimento possa ser aplicado adequadamente.

O semivariograma permite representar a variação quantitativa das variáveis regionalizadas através do cálculo da semivariância segundo a distância entre pontos, medindo a dissimilaridade entre os dados. Quanto menor o valor da semivariância, mais parecidos são os dados entre os



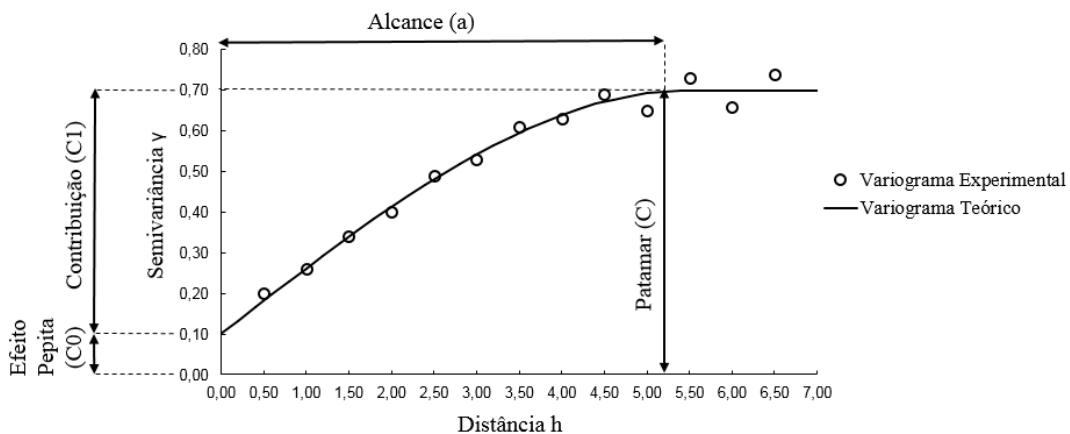
pontos analisados. Portanto, se há dependência espacial é natural que o valor da semivariância cresça com a distância e se estabilize em um patamar, a partir do qual não há mais dependência espacial. O semivariograma é dado pela Equação 4.

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [Z(x+h) - Z(x)]^2 \quad (4)$$

Em que:  $\gamma(h)$  função semivariograma em  $h$ ;  $h$  distância entre variáveis;  $Z(x)$  variável aleatória  $Z$  em  $x$ ;  $n$  número de elementos contabilizados.

Para a elaboração de semivariograma, é necessário determinar os valores de ângulo (direção), distância entre pontos ( $h$ ) e as tolerâncias do passo (distâncias ou lags) e angular. Os pares de pontos utilizados para o cálculo incluem o ponto de origem do vetor e os pontos dentro da área delimitada pelas tolerâncias. Todos os pontos são analisados para diferentes distâncias, múltiplas de  $h$ , de forma a ser possível construir um gráfico relacionando variância  $\gamma$  e distância entre os pares ( $h$ ), como apresentado na Figura 1.

Os pontos notáveis do semivariograma são: alcance ( $a$ ), corresponde à distância até a qual há correlação espacial; patamar ( $C$ ), corresponde ao valor da semivariância no alcance; efeito pepita ( $C_0$ ), efeito observado devido às descontinuidades do semivariograma para pequenas distâncias; contribuição ( $C_1$ ), diferença entre o patamar e o efeito pepita, corresponde à contribuição do modelo teórico.



**Figura 1:** Esquema semivariograma teórico. Adaptado de Druck *et al.* (2004)

A modelagem dos semivariogramas experimentais é realizada a partir da análise do melhor modelo teórico que se ajusta aos pontos do semivariograma experimental. Os modelos mais usuais são: efeito pepita puro (patamar constante indicando ausência de autocorrelação espacial), esférico, exponencial e gaussiano.

A função semivariograma pode demonstrar uma estrutura espacial isotrópica, com semivariogramas com distribuições similares em todas as direções, ou anisotrópica, apresentando distribuições diferentes para diferentes direções.

### 3. MATERIAIS E MÉTODO

#### 3.1. Materiais

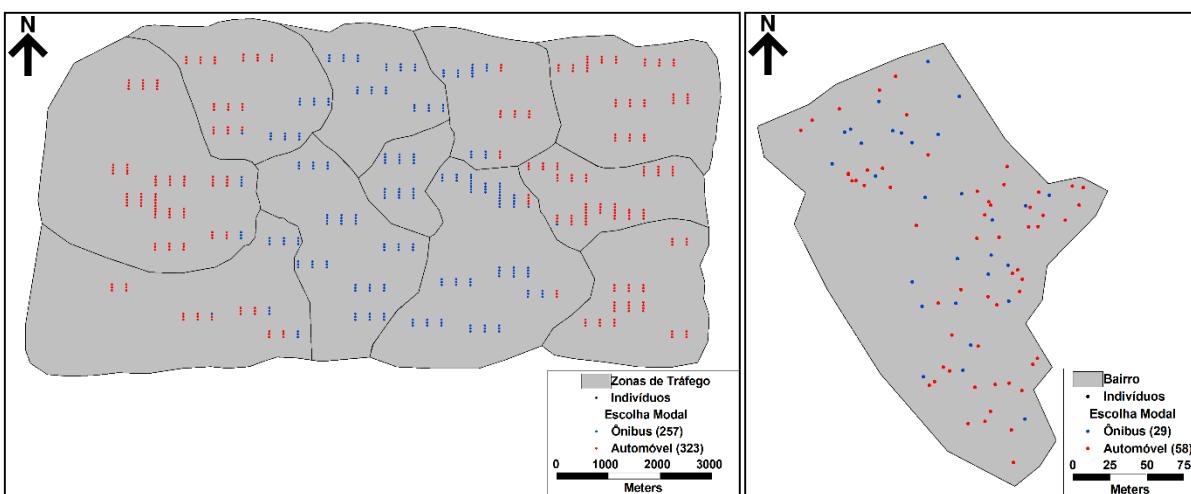
Os bancos de dados utilizados no presente projeto compreendem dados desagregados por indivíduos, são eles: uma cidade fictícia, construída de modo a apresentar alta correlação espacial (Assirati *et al.*, 2016), e o bairro Santa Felícia da cidade de São Carlos – SP (Rodrigues da Silva, 2008).

O banco de dados da cidade fictícia é proveniente do trabalho de Assirati *et al.* (2016), o qual



consiste numa população sintética com valores de variáveis espacialmente correlacionados. Tais dados foram obtidos, anteriormente, através do procedimento de interpolação bi-linear. A variável escolha modal é qualitativa binária, caso o modo escolhido seja ônibus (0) ou automóvel (1), como apresentado na Figura 2. A variável possui média de 0,557 e variância de 0,247. A amostra é composta por 580 observações.

O banco de dados do bairro Santa Felícia da cidade de São Carlos foi obtido através de um recorte da pesquisa Origem Destino, realizada na cidade em 2007/2008. Na pesquisa constam os modos utilizados pelos indivíduos entrevistados, a codificação utilizada indica modo ônibus (0) e modo automóvel (1), como apresentado na Figura 2. A variável possui média de 0,667 e variância de 0,225. A amostra do estudo de caso real é composta por 87 observações.



**Figura 2:** Representação da variável Escolha Modal, onde 0 corresponde ao modo ônibus e 1 corresponde ao modo automóvel: a) à esquerda, cidade fictícia produzida por (Assirati *et al.*, 2016), b) à direita, bairro Santa Felícia, São Carlos (Rodrigues da Silva, 2008).

Nesse trabalho foram utilizados aplicativos como *IBM SPSS Statistics 22*, *TransCAD Transportation Planning Software – Caliper Corporation*, *GeoDA*, *SGeMS – Stanford Geoestatistical Modeling Software*.

### 3.1. Método

O indicador global de associação espacial, baseado em conceitos variográficos, segue as seguintes premissas: (1) Para cada vizinhança de distância pré-definida é obtido um valor de semivariância; (2) Este valor é avaliado de forma a verificar se existe autocorrelação espacial nos dados analisados; (3) O indicador  $L_i$  (Equação 5) é dado em função dos valores de uma determinada variável na localização  $i$  e dos valores dos vizinhos a  $i$  para essa variável; (4) A vizinhança é determinada através do parâmetro de suavização utilizado na técnica de Regressão Geograficamente Ponderada.

$$L_i = f(y_i, y_{J_i}) \quad (5)$$

Em que:  $L_i$  indicador;  $y_i$  variável em  $i$ ;  $y_{J_i}$  valores dos vizinhos para variável  $i$ .

O indicador desenvolvido deve ser tal que permita inferir a significância estatística do padrão espacial da localização  $i$ . Isso requer que a probabilidade que o indicador seja superior ao valor crítico deva ser inferior ao nível de significância utilizado.

$$\text{Prob}[L_i > \delta_i] \leq \alpha_i \quad (6)$$

Em que:  $L_i$  indicador;  $\delta_i$  valor crítico;  $\alpha_i$  significância escolhida ou nível de pseuso-significância, por exemplo, resultado de um teste aleatório.



As etapas do procedimento metodológico proposto são descritas nas subseções subsequentes.

### *3.1.1. Obtenção da Probabilidade de Escolha Modal*

A probabilidade de escolha modal foi determinada aplicando-se Redes Neurais Artificiais (RNAs) e utilizando as demais variáveis do banco de dados original como variáveis independentes. A partir das variáveis originais, é obtida a variável probabilidade de escolha modal. O objetivo desta etapa foi obter uma variável contínua variando entre zero e um para a proposta do indicador global. Vale ressaltar que a ideia futura é tornar o indicador adequado para qualquer tipo de variável.

### *3.1.2. Determinação da primeira vizinhança*

Para determinar a distância que compreende a vizinhança ideal foi utilizado o conceito de parâmetro de suavização (*bandwidth*), proveniente da técnica de Regressão Geograficamente Ponderada (RGP). A técnica é utilizada para modelar fenômenos espaciais não estacionários de maneira local, ajustando um modelo de regressão para cada ponto utilizando como entrada os parâmetros dos pontos vizinhos. Os pontos de interesse para o modelo são determinados a partir do parâmetro de suavização (Fotheringham *et al.*, 2002).

É necessário determinar critérios de seleção do parâmetro de suavização, de forma a encontrar um valor médio adequado. Os critérios de seleção mais utilizados foram as minimizações através da validação cruzada (CV) e do critério de informação Akaike (AIC), os quais geralmente não apresentam diferenças significativas entre si (Fotheringham *et al.*, 2002).

### *3.1.3. Análise Variográfica*

A análise variográfica da variável em análise inicia-se através da definição dos parâmetros para cálculo dos semivariogramas experimentais. São definidos os parâmetros de lag (distância entre pontos, obtida através do parâmetro de suavização da técnica RGP), tolerância de lag (metade do lag utilizado), tolerância angular, tolerância do passo e direções nas quais serão calculadas os semivariogramas (principal e ortogonal). Com os valores escolhidos para lag e tolerância de lag, são determinados indicadores para faixas ou intervalos de distâncias.

Assim, os semivariogramas experimentais são calculados e analisados, identificando-se as direções principal (de máxima continuidade, menor variação de valores) e ortogonal (perpendicular a direção principal, de maior variação de valores).

Posteriormente, o semivariograma teórico é ajustado para todas as direções ao mesmo tempo, utilizando o mesmo efeito pepita, número de estruturas e tipo de função. Se a estrutura espacial é isotrópica a mesma função semivariograma irá se ajustar a todas as direções. Caso contrário há anisotropia, o alcance é dado por um elipsoide definido pelos coeficientes de alcance  $a$ . Os valores pontuais obtidos em ambos os semivariogramas experimentais e teóricos são os indicadores globais por vizinhança. Para determinar se esses valores são produto de um fenômeno espacialmente correlacionado ou de uma distribuição aleatória qualquer é desenvolvido um teste de hipótese.

### *3.1.4. Elaboração do Teste de Hipótese*

O teste de hipótese utilizado corresponde a um teste de pseudo-significância, assim como aplicado nos índices de Moran (1950). Para estimar a significância do índice proposto é necessário associar este a uma distribuição de probabilidades conhecida, como a distribuição normal. As hipóteses básicas do teste são formuladas como: (1) *Hipótese Nula*: A variável é espacialmente aleatória; (2) *Hipótese Alternativa*: A variável é espacialmente dependente, existe um padrão espacial.



Assim, gerou-se um banco de dados com as mesmas coordenadas de longitude e latitude que o banco de dados original, porém a variável de probabilidade de escolha modal foi substituída por valores uniformemente aleatórios entre 0 e 1. Em seguida, foi calculado e ajustado um semivariograma experimental omnidirecional e um semivariograma teórico para os dados simulados.

Como os dados gerados são aleatórios, não é possível observar dependência espacial. Quando isto ocorre o semivariograma tende a valores correspondentes a semivariância total dos dados, formando uma reta horizontal próximo a esse valor. Nessa situação, o gráfico é interpretado como efeito pepita puro.

Esse procedimento foi repetido 100 vezes, as semivariâncias teóricas e experimentais foram analisadas em conjunto para cada vizinhança. Através do teste de *Kolmogorov-Smirnov*, verificou-se se a distribuição de valores pontuais dos semivariogramas experimentais e teóricos correspondia a uma distribuição normal para todas as vizinhanças analisadas.

Sendo assim, é possível encontrar os valores críticos correspondentes para o nível de significância desejada. Os pontos críticos são obtidos através do cálculo que relaciona a distribuição produzida com a distribuição normal padrão.

### *3.1.5. Teste de Significância*

Obtendo-se os valores dos pontos críticos, realiza-se um teste de significância, avaliando se através dos valores de semivariância encontrado nos dados experimentais e teóricos por vizinhança, na região a ser analisada, aceita-se ou não a hipótese nula de aleatoriedade.

### *3.1.6. Comparação com indicadores globais comumente utilizados na literatura*

Finalmente, calcula-se o índice I de Moran para o conjunto de dados analisados. O valor obtido deve ser comparado aos valores do indicador global proposto de forma a validar o procedimento metodológico aqui proposto.

## **4. RESULTADOS E DISCUSSÕES**

### **4.1. Cidade Fictícia**

#### *4.1.1. Obtenção da Probabilidade de Escolha Modal*

O processamento da Rede Neural Artificial foi realizado através do software IBM SPSS Statistics 22. A variável probabilidade de escolha modal foi obtida através de um modelo com as seguintes variáveis: distância, tempo de viagem de ônibus, tempo de viagem de automóvel, preço do modo ônibus e preço do modo automóvel. Com média de 0,556 e variância de 0,209, sua distribuição é bastante próxima da variável escolha modal original.

#### *4.1.2. Determinação da primeira vizinhança*

Para determinação do parâmetro de suavização, foi utilizado o critério de seleção de validação cruzada e a função de ponderação gaussiana fixa. Através do software R, obteve-se o parâmetro de suavização ideal através do critério de seleção de validação cruzada igual a 0,506 km e através do critério AIC igual a 0,629 km. Observou-se resultados bastante similares para análise variográfica. Adotou-se o valor de 0,506 km para o lag ou vizinhanças. Assim, a primeira vizinhança corresponde aos pontos situados ao intervalo de distância entre 0 e 0,506km. A opção pelo uso de intervalos de distâncias justifica-se pela necessidade de obtenção de um número maior de pares de pontos.

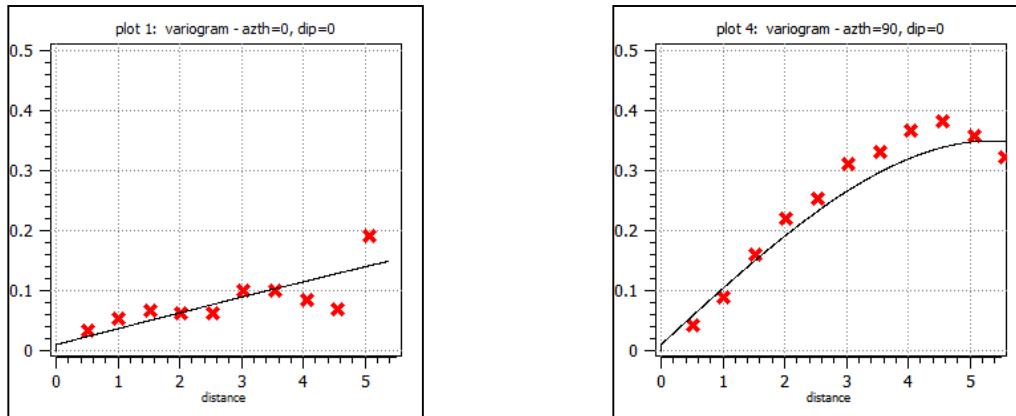
#### *4.1.3. Análise Variográfica*

A análise variográfica da variável probabilidade de escolha modal foi realizada para as direções 0°, 30°, 60°, 90°, 120° e 150°. A distância máxima de análise corresponde a 5,50 km, a partir



da qual não é observada mais dependência espacial. Os parâmetros foram fixados da seguinte forma: lag de 0,506 km (valor obtido na etapa anterior), tolerância do lag de 0,253 km (metade do lag), tolerância angular de 22°, tolerância do passo de 0,253 km (metade do lag). Obtém-se que a direção principal corresponde a 0° e a direção ortogonal a 90°.

O semivariograma teórico foi ajustado para todas as direções ao mesmo tempo utilizando uma função esférica, o efeito pepita ( $C_0$ ) corresponde a 0,010 e a contribuição ( $C_1$ ) a 0,340. Observou-se anisotropia. Os variogramas são apresentados na Figura 3 e os valores de semivariâncias experimentais e teóricas para as direções principal e ortogonal das primeiras vizinhanças são apresentados na Tabela 2.



**Figura 3:** Semivariogramas experimentais e teóricos para a direção a) principal e b) ortogonal da variável probabilidade de escolha modal – Cidade Fictícia

**Tabela 2:** Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal para as primeiras vizinhanças – Cidade Fictícia

Vizinhança	Distância (h)	Semivariância ( $\gamma$ ) Experimental		Semivariância ( $\gamma$ ) Teórica	
		Dir. Princ. (0°)	Dir. Ortog. (90°)	Dir. Princ. (0°)	Dir. Ortog. (90°)
1	0,506	0,032	0,043	0,023	0,058
2	1,013	0,053	0,090	0,037	0,105
3	1,519	0,066	0,160	0,050	0,150
4	2,026	0,064	0,220	0,063	0,192
5	2,532	0,061	0,254	0,076	0,232

#### 4.1.4. Elaboração do Teste de Hipótese

Utilizando o software R foi gerado um banco de dados com as mesmas coordenadas que o banco de dados original da cidade fictícia. Os valores adotados para a variável probabilidade de escolha modal correspondem a números aleatórios entre zero e um.

Para os dados aleatórios, os semivariogramas desenvolvidos são ominidirecionais, de forma a considerar quaisquer tendências espaciais que os novos dados possam apresentar. Os dados aleatórios não possuem estrutura espacial, produziram semivariogramas com tendência constante, bem representados pelo modelo efeito pepita puro. Assim, esse procedimento foi repetido 100 vezes, gerando novos dados aleatórios a cada repetição. As semivariâncias experimentais e teóricas obtidas foram analisadas por vizinhança, formando um conjunto de 100 observações por vizinhança. Constatou-se que todas possuíam resultados muito similares de média e desvio padrão. Assim, optou-se por agregá-las para as dez primeiras vizinhanças, formando um conjunto de 1000 observações. Aplicou-se o teste de normalidade *Kolmogorov-Smirnov* ao novo conjunto de dados, comprovando que a distribuição é similar a normal.



Adotando a distribuição dos dados como normal, é possível obter a probabilidade de um certo evento ocorrer por meio da área sob a curva normal padrão. Para isso é definido um valor z padronizado para análise bilateral da curva, definindo o grau de significância desejado. Com o valor de z, a média e desvio padrão da amostra analisada é possível obter os valores críticos correspondentes para o conjunto de dados em questão.

Dessa forma, se o valor de semivariância obtido estiver entre os limites estabelecidos à hipótese nula deve ser retida com um dado nível de significância. Caso o valor esteja fora desses limites, a hipótese nula é rejeitada e assume-se que há um padrão espacial. A Tabela 3 apresenta os valores críticos por significância desejada, tendo como base os valores de semivariância obtidos através da análise de dados aleatórios.

**Tabela 3:** Valores críticos por significância desejada para os valores de semivariância

Intervalo de Confiança	z	Significância (p-valor αi)	Limite inferior	Limite superior
80%	$\pm 1,280$	0,200	0,078	0,088
85%	$\pm 1,440$	0,150	0,078	0,089
90%	$\pm 1,645$	0,100	0,077	0,090
95%	$\pm 1,960$	0,050	0,075	0,091
99%	$\pm 2,575$	0,010	0,073	0,093

#### 4.1.5. Teste de Significância

Utilizando a Tabela 3 de valores críticos por significância desejada, é possível analisar os dados obtidos na Tabela 2, que apresenta os resultados de semivariância experimentais e teóricos para direção principal e ortogonal.

Os pontos de maior interesse possuem semivariância abaixo do limite inferior, observa-se que para 99% de significância na direção principal, pode se considerar até a 5ª vizinhança experimental e até a 4ª vizinhança teórica como existência de correlação espacial dos dados. Para a direção ortogonal os valores de semivariância são 99% significativos apenas para a primeira vizinhança experimental e teórica. No entanto, a ideia aqui proposta é que o indicador seja associado à direção principal apenas.

#### 4.1.6. Comparação com indicadores globais comumente utilizados na literatura

O valor do índice I de Moran para os dados de probabilidade de escolha modal para a cidade fictícia corresponde à aproximadamente 0,824 para a primeira vizinhança, como esperado, reflete que alta correlação positiva entre os dados. A Tabela 4 traz uma síntese do cálculo do indicador de Moran, em comparação ao indicador proposto até a quinta vizinhança. O indicador proposto corresponde aos resultados do semivariograma teórico. Os resultados do semivariograma experimental não foram utilizados pois apresentam variações abruptas entre as vizinhanças.

**Tabela 4:** Comparação de valores do indicador, z e status da hipótese nula (para significância de 95%) para índice I de Moran e indicador proposto (direção principal) – Cidade Fictícia

Vizinhança	Distância (h)	Moran			Indicador Proposto				
		Índice	z	p-valor	Hipótese Nula	Índice	z	p-valor	Hipótese Nula
1	0,506	0,824	43,658	0,000	Rejeita	0,023	-15,377	0,000	Rejeita
2	1,013	0,765	73,972	0,000	Rejeita	0,037	-11,798	0,000	Rejeita
3	1,519	0,632	84,624	0,000	Rejeita	0,050	-8,476	0,000	Rejeita
4	2,026	0,514	99,497	0,000	Rejeita	0,063	-5,153	0,000	Rejeita
5	2,532	0,401	112,510	0,000	Rejeita	0,076	-1,830	0,034	Aceita



#### 4.2. Bairro Santa Felícia, São Carlos – SP

##### 4.2.1. Obtenção da Probabilidade de Escolha Modal

A variável probabilidade de escolha modal foi obtida através de um modelo de RNAs com as seguintes variáveis: posse de motocicleta, número de automóveis, posse de CNH, se o indivíduo é estudante, número de viagens realizadas por dia e se o indivíduo é maior de 52 anos. Com média de 0,679 e variância de 0,077, sua distribuição é bastante próxima da variável escolha modal original.

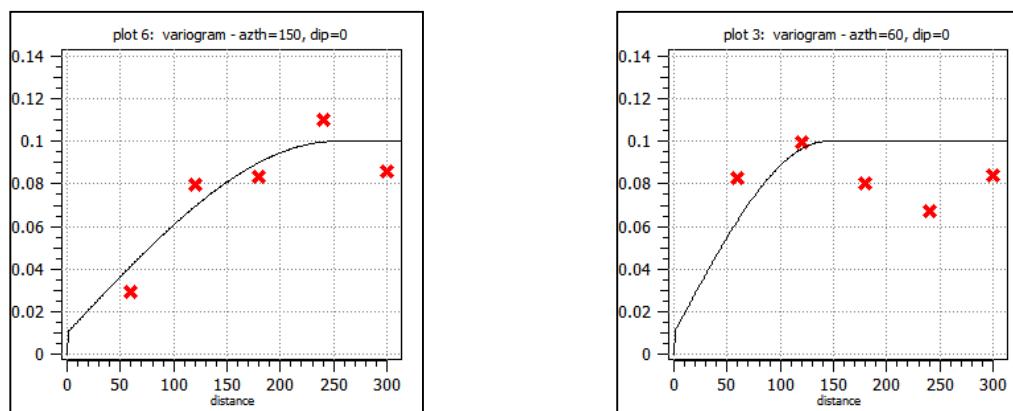
##### 4.2.2. Determinação da primeira vizinhança

Através do software R, obteve-se o parâmetro de suavização ideal através do critério de seleção de validação cruzada e do critério AIC igual a 1.250,644 m. Esse valor corresponde à cerca de metade a distância máxima entre os dados, não sendo adequado para o estudo de caso. Optou-se, então, por utilizar o valor de 60 m, o qual foi utilizado para o desenvolvimento de alguns semivariogramas e mostrou-se satisfatório.

##### 4.2.3. Análise Variográfica

A análise variográfica da variável probabilidade de escolha modal foi realizada para as direções 0°, 30°, 60°, 90°, 120° e 150°. A distância máxima de análise corresponde a 300 m, a partir da qual não é observada mais dependência espacial. Os parâmetros foram fixados da seguinte forma: lag de 60 m, tolerância do lag de 30 m, tolerância angular de 22°, tolerância do passo de 30 m. Observando os semivariogramas experimentais, obtém-se que a direção principal corresponde a 150° e a direção ortogonal a 60°.

O semivariograma teórico foi ajustado para todas as direções ao mesmo tempo utilizando uma função esférica, o efeito pepita ( $C_0$ ) corresponde a 0,010 e a contribuição ( $C_1$ ) a 0,090. Os semivariogramas são apresentados na Figura 4, os valores de semivariâncias correspondentes para as primeiras vizinhanças são apresentados na Tabela 5.



**Figura 4:** Semivariogramas experimentais e teóricos para a direção a) principal e b) ortogonal da variável probabilidade de escolha modal – Bairro Santa Felícia, São Carlos – SP

**Tabela 5:** Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal para as primeiras vizinhanças – Bairro Santa Felícia, São Carlos – SP

Vizinha- nça	Distân- cia (h)	Semivariância ( $\gamma$ ) Experimental		Semivariância ( $\gamma$ ) Teórico	
		Dir. Principal (150°)	Dir. Ortogonal (60°)	Dir. Principal (150°)	Dir. Ortogonal (60°)
1	60,000	0,029	0,083	0,042	0,063
2	120,000	0,080	0,099	0,069	0,096
3	180,000	0,083	0,080	0,087	0,096
4	240,000	0,110	0,067	0,087	0,096



#### 4.2.4. Elaboração do Teste de Hipótese

Como o teste de hipótese desenvolvido em 4.1.4 utiliza dados aleatórios, é válido utilizar os valores críticos, obtidos anteriormente, para a análise do bairro Santa Felícia, São Carlos - SP.

#### 4.2.5. Teste de Significância

Utilizando a Tabela 3 de valores críticos por significância desejada, é possível analisar os dados obtidos na Tabela 5, que apresenta os resultados de semivariância experimentais e teóricos para direção principal e ortogonal.

Os pontos de maior interesse possuem semivariância abaixo do limite inferior, observa-se que para 99% de significância na direção principal pode se considerar até a primeira vizinhança experimental e até a 2<sup>a</sup> vizinhança teórica. Para a direção ortogonal os valores de semivariância são 99% significativos apenas para a primeira vizinhança teórica.

#### 4.2.6. Comparação com indicadores globais comumente utilizados na literatura

O valor do índice I de Moran para os dados de probabilidade de escolha modal para o bairro Santa Felícia, São Carlos – SP corresponde à aproximadamente -0,057, para primeira vizinhança, refletindo baixa correlação espacial negativa. Apesar de valor baixo do I de Moran mesmo para a primeira vizinhança, o indicador proposto neste trabalho atestou correlação espacial na 1<sup>a</sup> vizinhança. A Tabela 6 sintetiza a análise comparativo entre o Índice de Moran e o Indicador proposta até a quarta vizinhança. Assim como na aplicação no banco de dados da Cidade Fictícia, utilizaram-se os dados do semivariograma teórico para a proposta de indicador.

**Tabela 6:** Comparação de valores do indicador, z e status da hipótese nula (para significância de 95%) para índice I de Moran e indicador proposto (direção principal) – Santa Felícia, São Carlos – SP

Vizinhança	Distância (h)	Moran				Indicador Proposto			
		Índice	z	p-valor	Hipótese Nula	Índice	z	p-valor	Hipótese Nula
1	60	-0,057	-0,697	0,253	Aceita	0,042	-10,520	0,000	Rejeita
2	120	-0,158	-1,495	0,071	Aceita	0,069	-3,619	0,000	Rejeita
3	180	-0,067	-0,632	0,273	Aceita	0,087	0,981	0,164	Aceita
4	240	-0,051	-0,570	0,282	Aceita	0,087	0,981	0,164	Aceita

## 5. CONCLUSÕES

O presente trabalho propõe um procedimento metodológico para o cálculo de um indicador de associação espacial global, o qual permite quantificar a dependência espacial de variáveis de transporte, como a escolha modal dos usuários. O indicador proposto baseia-se no conceito de semivariograma proveniente da geoestatística. O gráfico de semivariograma permite a análise visual da estrutura espacial da variável, identificando direção principal, bem como viabiliza o uso de variáveis discretas e binárias.

Desta forma, os valores pontuais dos semivariogramas experimentais e teóricos, em diferentes vizinhanças, são os indicadores aqui propostos. Observa-se que o indicador proveniente do semivariograma teórico apresenta um comportamento mais contínuo e adequado à aplicação, enquanto o semivariograma experimental apresenta variações abruptas entre os resultados obtidos. Tais indicadores são associados a um teste de hipótese, fazendo-se uma comparação com dados espacialmente aleatórios. A validação do indicador proposto é realizada após a comparação com indicadores comumente utilizados na literatura, como o I de Moran. Observou-se que valores altos de I de Moran estão associados às vizinhanças onde a hipótese de aleatoriedade espacial aqui proposta foi refutada.



O índice  $c$  de Geary (1954) assemelha-se ao cálculo realizado no presente trabalho, porém não considera possíveis anisotropias entre diferentes direções, realizando um cálculo omnidirecional. Da mesma forma que a estatística G(d) de Getis e Ord (1992), o indicador desenvolvido representa um índice de concentração espacial, pois identifica apenas autocorrelação espacial positiva.

Além disso, é importante destacar que o método desenvolvido é aplicável a dados binários e contínuos entre zero e um. Para uma aplicação mais ampla ainda são necessários testes com variáveis quantitativas padronizadas.

#### Agradecimentos

O presente trabalho foi realizado com apoio do CNPq.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Assirati, L.; Rocha, S. S.; Caldas, M. U.; Pitombo, C. S. (2016) Interpolação bi-linear para simulação de dados espacialmente correlacionados: uma aplicação relativa à demanda por transportes. In: ANPET 2016, Rio de Janeiro. *Anais...* Rio de Janeiro.
- Cliff, A.; Ord, K. (1972) Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*. v. 4, p. 267-284 DOI: 10.1111/j.1538-4632.1972.tb00475.x
- Druck, S.; Carvalho, M. S.; Câmara, G.; Monteiro, A.V.M. (2004) *Análise Espacial de Dados Geográficos*. Brasília, EMBRAPA, 2004 (ASBN:85-7383-260-6) 209 p.
- Fortin, M. J.; Drapeau, P.; Legendre, P. (1989) Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*. v. 83, p. 209-222 DOI: 10.1007/BF00031693
- Fotheringham, A. S.; Brunsdon, C.; Charlton, M. (2002) Geographically weighted regression: the analysis of spatially varying relationships. *John Wiley & Sons*, 284 p.
- Geary, R.C. (1954) The contiguity ratio and statistical mapping. *The Incorporated Statistician*. v. 5, p. 115-145 DOI: 10.2307/2986645
- Getis, A.; Ord, J. K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*. v.24, p. 189-206 DOI: 10.1111/j.1538-4632.1992.tb00261.x
- Loureiro, C. F. G.; Silva, H. N.; Carvalho L. E. X. (2006) Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. Panorama Nacional de Pesquisa em Transportes. In: ANPET, 20, Curitiba. *Anais...* Curitiba, v.1, p.479-491.
- Matheron, G. (1971) *La Théorie des Variables Régionalisées et ses Applications*, Les Cahiers de Morphologie Mathématique de Fontainebleau. École Nationale Supérieure des Mines de Paris, Paris, nº5 211 p.
- Moran, P. A. (1950) Notes on continuous stochastic phenomena. *Biometrika*. v. 37, p. 17-23 DOI:10.2307/2332142
- Rodrigues da Silva, A. N. R. (2008) *Pesquisa Origem Destino da Cidade de São Carlos*. Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos – USP. São Carlos.
- Sener, I.N.; Pendyala, R.M.; Bhat, C.R. (2010) Accommodating spatial correlation across choice alternatives in discrete choice models: An application to modeling residential location choice behavior. *Journal of Transport Geography*. v. 19, p. 294-303 DOI: 10.1016/j.jtrangeo.2010.03.013
- Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* v. 46, p. 234-240 DOI: 10.2307/143141