UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

CLÁUDIA CRISTINA BAPTISTA RAMOS NAIZER

Procedimento Metodológico para Proposta de Indicadores de Associação
espacial Global e Local através de conceitos variográficos

São Carlos

2018

CLÁUDIA CRISTINA BAPTISTA RAMOS NAIZER

**Procedimento Metodológico para Proposta de Indicadores de Associação
espacial Global e Local através de conceitos variográficos**

Versão Original

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Mestre em Ciências, Programa de Pós-Graduação em Engenharia de Transportes.

Área de concentração: Planejamento e Operação de Sistemas de Transporte.

Orientadora: Profa. Dra. Cira Souza Pitombo

São Carlos

2018

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

N161p	<p>Naizer, Cláudia Cristina Baptista Ramos</p> <p>Procedimento Metodológico para Proposta de Indicadores de Associação espacial Global e Local através de conceitos variográficos / Cláudia Cristina Baptista Ramos Naizer; orientadora Cira Souza Pitombo. São Carlos, 2018.</p> <p>Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2018.</p> <p>1. Variograma. 2. Índice de Moran. 3. Análise Exploratória de Dados Espaciais. 4. Geoestatística. I. Título.</p>
-------	--

FOLHA DE JULGAMENTO

Candidata: Engenheira **CLAUDIA CRISTINA BAPTISTA RAMOS NAIZER**.

Título da dissertação: "Procedimento metodológico para proposta de indicadores de associação espacial global e local através de conceitos variográficos".

Data da defesa: 16/03/2018.

Comissão Julgadora:

Resultado:

Profa. Associada **Cira Souza Pitombo**
(Orientadora)
(Escola de Engenharia de São Carlos/EESC)

Aprovada

Prof. Dr. **Gustavo Garcia Manzato**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP-Bauru)

Aprovada

Profa. Dra. **Flavia da Fonseca Feitosa**
(Universidade Federal do ABC/UFABC)

Aprovada

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:
Profª. Associada **Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luis Fernando Costa Alberto**

AGRADECIMENTOS

Agradeço aos meus pais, Elizabeth e Claudi, pelos ensinamentos de perseverança, coragem e amor. Obrigada a Cristina, Rosângela e Gil por manterem a família unida enquanto estive longe e me receberem sempre com carinho incondicional. Agradeço aos meus avós Arnaldo e Antonieta que sempre acreditaram no meu potencial.

Obrigada à professora Cira Pitombo pela confiança e pela orientação, estando sempre disponível e empolgada com o desenvolvimento do projeto. Também agradeço ao professor Antônio Néelson Rodrigues da Silva pelas contribuições para o desenvolvimento desse trabalho.

Agradeço à professora Ana Paula Larocca pelo zelo com o Programa de Engenharia de Transportes da EESC-USP e pela parceria formada ao longo do mestrado, tornando minha gestão, como representante discente, mais enriquecedora.

Obrigada pela amizade e o apoio durante a pós-graduação de Francine Tan, Bruna Kuramoto, Denise Capasso, Julia Simoni, Adrian Grasson Filho, Luan Carvalho, Adriano Felicio e Cintia Campos. Pelo compartilhamento de ideias e conselhos, agradeço a Lucas Assirati e Tiago Torquato. Obrigada José Venâncio por ser meu vice-representante e dividir comigo responsabilidades e conquistas.

Muito obrigada à família Grupo de Som, principalmente a Henrique Leal, Caio Brandão, Lucas Lucafó e Nathália Felipe, presentes nas horas de descontração e desabafos, sendo sempre meu porto seguro. Obrigada aos amigos Michel Bessani, Viviane Kim, Rafael Mori e Fernando Stroppa pelo suporte e trocas de conhecimentos.

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro ao longo do desenvolvimento do presente projeto.

RESUMO

NAIZER, C. C. B. R. **Procedimento Metodológico para Proposta de Indicadores de Associação espacial Global e Local através de conceitos variográficos**. 2018. 198 p. Dissertação (Mestrado em Ciências) – Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Entre as técnicas de análise exploratória de dados espaciais encontram-se os indicadores de associação espacial, que mensuram o grau de dependência espacial dos dados analisados e são aplicáveis apenas a dados quantitativos. Outro procedimento disponível é a geoestatística, a qual se baseia no variograma, descrevendo quantitativa e qualitativamente a estrutura espacial de determinada variável. Neste trabalho, utilizam-se conceitos do variograma para desenvolver um indicador de associação espacial global (SIVAR-G) e um indicador de associação espacial local (SIVAR-L). São utilizados dois bancos de dados: dados binários de escolha modal de uma Cidade Fictícia e dados de média de viagens por modo automóvel por domicílio para um recorte da região central da cidade de São Paulo (Pesquisa de Mobilidade 2012). Em ambos os casos, o indicador global, para diferentes vizinhanças, foi calculado com base em valores padronizados, provenientes do variograma experimental e teórico. Em seguida, aplicou-se um teste de hipótese baseado em pseudo-significância para avaliar a significância do indicador proposto previamente. Por fim, os resultados do indicador proposto foram comparados ao índice de Moran, calculado com os mesmos parâmetros. Para o indicador local, foi elaborado um procedimento similar, porém os cálculos foram feitos pontualmente. Cada observação do banco de dados teve um variograma experimental calculado e um variograma teórico modelado para uma análise omnidirecional. Um teste de hipótese similar ao do indicador global foi desenvolvido e aplicado. Assim obtiveram-se indicadores de associação espacial local ponto a ponto. Conclui-se que o indicador SIVAR-G possui desempenho satisfatório na estimação de associação espacial para dados contínuos e binários, mostrando-se sensível a anisotropia dos dados. O indicador SIVAR-L é capaz de identificar “bolsões” de associação espacial. É aplicável a dados contínuos e binários. Os indicadores propostos permitem a modelagem de variogramas teóricos globais e locais, fornecendo uma maior riqueza de detalhes da estrutura espacial dos dados. Os indicadores SIVAR baseiam-se na dissimilaridade espacial, enquanto o índice de Moran e LISA baseiam-se na similaridade espacial.

Palavras-chave: Variograma, Índice de Moran, LISA, Análise Exploratória de Dados Espaciais, Geoestatística

ABSTRACT

NAIZER, C. C. B. R. **Methodological Procedure for proposal of global and local indicators of spatial association based on variographic concepts.** 2018. 198 f. Thesis (Master of Science) –School of Engineering of São Carlos, University of São Paulo, São Carlos, 2018.

Among the exploratory spatial data analysis tools, there are the indicators of spatial association, which measure the degree of spatial dependence of the analyzed data and can be applied to quantitative data. Other procedure available is the geostatistics, which is based on the variogram, describing quantitatively and qualitatively the spatial structure of a variable. The aim of this thesis is to use the concept of the variogram to develop a global indicator of spatial association (SIVAR-G) and a local indicator of spatial association (SIVAR-L). Two data bases were used: binary data of travel mode choice of a hypothetical city and mean of automobile trips by household to a region of São Paulos's center (Mobility Survey, 2012). In both cases, the global indicator, for different neighborhoods, was calculated based on standardized values, derived from the experimental and theoretical variogram. Then, a pseudo-significance test was applied to evaluate the significance of the previously proposed indicator. The results of the proposed indicator were compared to Moran's I, calculated with the same parameters. For the local indicator, it was made a similar procedure, however the calculation was punctual. For each observation of the database, it was calculated experimental variogram and adjusted a theoretical variogram for a omnidirectional analysis. A hypothesis test similar to the one applied in the global indicator was developed and applied. Therefore, it was obtained local indicators point by point. It was concluded that the indicator SIVAR-G has a satisfactory performance for spatial association of binary and continuous data, with sensibility for anisotropy cases. The SIVAR-L indicator is able to identify spatial association pockets and outliers. The local indicator is suitable for continuous and binary data. The developed indicators allow the modeling of theoretical global and local variograms, providing more details of the spatial structure of the data. The SIVAR indicators are based on spatial dissimilarity, while the Moran and LISA index are based on spatial similarity.

Key-words: Variogram, Moran's I, LISA, Exploratory Spatial Data Analysis, Geostatistics

LISTA DE ILUSTRAÇÕES

Figura 2-1: Diagrama de Espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, censo de 1991 (CAMARGO, 1998).	37
Figura 3-1: Esquema de cálculo de semivariogramas experimentais para distribuições irregulares. Adaptado de Camargo (1998).	41
Figura 3-2: Esquema semivariograma teórico. Adaptado de Burrough (1986)	43
Figura 4-1: Representação da variável Escolha Modal na cidade fictícia, produzida por Assirati <i>et al.</i> (2016)	48
Figura 4-2: Número médio de viagens por modo automóvel por quadrícula de 500 por 500 m.	49
Figura 4-3: Número médio de viagens por modo automóvel, associado aos centroides das quadrículas de 500 por 500 m.	50
Figura 4-4: Fluxograma do método para o Indicador Global de Associação Espacial proposto	53
Figura 4-5: Alcances do elipsoide de anisotropia.....	57
Figura 5-1: Semivariogramas experimentais e teóricos para a direção principal e ortogonal – Cidade Fictícia.....	65
Figura 5-2: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Cidade Fictícia.....	68
Figura 5-3: Comparação entre: a) os índices de Moran, indicador SIVAR-G e exposição do limite crítico para o nível de confiança de 95%, b) além dos valores já apresentados são representados os valores de semivariância para a direção ortogonal, de forma a comparar os resultados na direção principal e ortogonal – Cidade Fictícia.....	72
Figura 5-4: Semivariogramas experimentais e teóricos para a direção principal e ortogonal – Centro, São Paulo – SP.....	75
Figura 5-5: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Centro, São Paulo – SP	77
Figura 5-6: Comparação entre os índices de Moran e indicador proposto e exposição do limite crítico para o nível de confiança de 95% – Centro, São Paulo – SP	80
Figura 6-1: Semivariância Experimental para a primeira vizinhança resultante de análise local – Cidade Fictícia.....	84

Figura 6-2: Modelos Teóricos – Cidade Fictícia	86
Figura 6-3: Correlação entre valores observados e estimados pelos modelos teóricos – Cidade Fictícia.....	86
Figura 6-4: Semivariância Teórica para a primeira vizinhança resultante de análise local – Cidade Fictícia	88
Figura 6-5: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Cidade Fictícia	90
Figura 6-6: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a primeira vizinhança resultante de análise local – Cidade Fictícia.....	91
Figura 6-7: Significância dos valores de SIVAR-L para a primeira vizinhança resultante – Cidade Fictícia	92
Figura 6-8: Correlação do Indicador LISA para a primeira vizinhança – Cidade Fictícia	93
Figura 6-9: Significância do Indicador LISA para a primeira vizinhança – Cidade Fictícia...	93
Figura 6-10: Mapa de <i>Clusters</i> LISA – Cidade Fictícia	94
Figura 6-11: Pontos selecionados para comparação – Cidade Fictícia.....	96
Figura 6-12: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 204 – Cidade Fictícia	96
Figura 6-13: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 283 – Cidade Fictícia	97
Figura 6-14: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95%– Ponto ID 364 – Cidade Fictícia	97
Figura 6-15: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 390 – Cidade Fictícia	98
Figura 6-16: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 416 – Cidade Fictícia	98
Figura 6-17: Semivariância Experimental para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP	100
Figura 6-18: Semivariância Experimental para a primeira vizinhança resultante de análise local, com as regiões de baixa associação espacial em destaque – Centro, São Paulo – SP	101

Figura 6-19: Modelos Teóricos – Centro, São Paulo – SP	102
Figura 6-20: Correlação entre valores observados e estimados pelos modelos teóricos – Centro, São Paulo – SP.....	102
Figura 6-21: Semivariância Teórica para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP.....	103
Figura 6-22: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Centro, SP – SP....	106
Figura 6-23: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP	107
Figura 6-24: Indicador SIVAR-L, para a primeira vizinhança resultante de análise local, com os “bolsões” de associação espacial em destaque – Centro, São Paulo – SP.....	108
Figura 6-25: Significância das Semivariâncias Teóricas Padronizadas para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP	109
Figura 6-26: Correlação do Indicador LISA para a primeira vizinhança – Centro, São Paulo – SP.....	110
Figura 6-27: Significância do Indicador LISA para a primeira vizinhança – Centro, São Paulo – SP.....	110
Figura 6-28: Mapa de <i>Clusters</i> LISA – Centro, São Paulo – SP.....	111
Figura 6-29: Pontos selecionados para comparação – Centro, São Paulo – SP	113
Figura 6-30: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 196 – Centro, São Paulo – SP	114
Figura 6-31: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 229 – Centro, São Paulo – SP	114
Figura 6-32: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 301 – Centro, São Paulo – SP	115
Figura 6-33: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 430 – Centro, São Paulo – SP	115
Figura 6-34: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 724 – Centro, São Paulo – SP	116

Figura 7-1: Configuração (1) de indicadores tradicionais e (2) do indicador proposto para uma análise omnidirecional, d corresponde à distância da primeira vizinhança e a tolerância de (2) é dada por $d/2$	121
Figura 9-1: Recorte realizado na região central da cidade de São Paulo	135
Figura 9-2: Macrozonas da Região Central de São Paulo selecionada para o estudo	136
Figura 9-3: Legenda correspondente às zonas na Figura 9-2.....	137
Figura 9-4: Variogramas experimentais da variável escolha modal – Cidade Fictícia.....	139
Figura 9-5: Semivariogramas experimentais da variável escolha modal – Centro, São Paulo – SP	140
Figura 9-6: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a segunda vizinhança resultante de análise local – Cidade Fictícia.....	162
Figura 9-7: Significância das Semivariâncias Teóricas Padronizadas para a segunda vizinhança resultante de análise local – Cidade Fictícia	162
Figura 9-8: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a terceira vizinhança resultante de análise local – Cidade Fictícia.....	163
Figura 9-9: Significância das Semivariâncias Teóricas Padronizadas para a terceira vizinhança resultante de análise local – Cidade Fictícia	163
Figura 9-10: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quarta vizinhança resultante de análise local – Cidade Fictícia.....	164
Figura 9-11: Significância das Semivariâncias Teóricas Padronizadas para a quarta vizinhança resultante de análise local – Cidade Fictícia	164
Figura 9-12: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quinta vizinhança resultante de análise local – Cidade Fictícia.....	165
Figura 9-13: Significância das Semivariâncias Teóricas Padronizadas para a quinta vizinhança resultante de análise local – Cidade Fictícia	165
Figura 9-14: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sexta vizinhança resultante de análise local – Cidade Fictícia.....	166
Figura 9-15: Significância das Semivariâncias Teóricas Padronizadas para a sexta vizinhança resultante de análise local – Cidade Fictícia	166
Figura 9-16: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sétima vizinhança resultante de análise local – Cidade Fictícia.....	167
Figura 9-17: Significância das Semivariâncias Teóricas Padronizadas para a sétima vizinhança resultante de análise local – Cidade Fictícia	167

Figura 9-18: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a oitava vizinhança resultante de análise local – Cidade Fictícia	168
Figura 9-19: Significância das Semivariâncias Teóricas Padronizadas para a oitava vizinhança resultante de análise local – Cidade Fictícia.....	168
Figura 9-20: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a nona vizinhança resultante de análise local – Cidade Fictícia	169
Figura 9-21: Significância das Semivariâncias Teóricas Padronizadas para a nona vizinhança resultante de análise local – Cidade Fictícia.....	169
Figura 9-22: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a décima vizinhança resultante de análise local – Cidade Fictícia	170
Figura 9-23: Significância das Semivariâncias Teóricas Padronizadas para a décima vizinhança resultante de análise local – Cidade Fictícia.....	170
Figura 9-24: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a segunda vizinhança resultante de análise local – Centro, São Paulo - SP.....	171
Figura 9-25: Significância das Semivariâncias Teóricas Padronizadas para a segunda vizinhança resultante de análise local – Centro, São Paulo - SP.....	171
Figura 9-26: : Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a terceira vizinhança resultante de análise local – Centro, São Paulo - SP.....	172
Figura 9-27: Significância das Semivariâncias Teóricas Padronizadas para a terceira vizinhança resultante de análise local – Centro, São Paulo - SP	172
Figura 9-28: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quarta vizinhança resultante de análise local – Centro, São Paulo - SP.....	173
Figura 9-29: Significância das Semivariâncias Teóricas Padronizadas para a quarta vizinhança resultante de análise local – Centro, São Paulo – SP.....	173
Figura 9-30: : Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quinta vizinhança resultante de análise local – Centro, São Paulo - SP.....	174
Figura 9-31: Significância das Semivariâncias Teóricas Padronizadas para a quinta vizinhança resultante de análise local – Centro, São Paulo - SP	174
Figura 9-32: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sexta vizinhança resultante de análise local – Centro, São Paulo - SP.....	175
Figura 9-33: Significância das Semivariâncias Teóricas Padronizadas para a sexta vizinhança resultante de análise local – Centro, São Paulo – SP.....	175
Figura 9-34: Correlação do Indicador LISA para a segunda vizinhança – Cidade Fictícia ...	176
Figura 9-35: Significância do Indicador LISA para a segunda vizinhança – Cidade Fictícia	176

Figura 9-36: Mapa de <i>Clusters</i> LISA para a segunda vizinhança – Cidade Fictícia	177
Figura 9-37: Correlação do Indicador LISA para a terceira vizinhança – Cidade Fictícia....	177
Figura 9-38: Significância do Indicador LISA para a terceira vizinhança – Cidade Fictícia	178
Figura 9-39: Mapa de <i>Clusters</i> LISA para a terceira vizinhança – Cidade Fictícia	178
Figura 9-40: Correlação do Indicador LISA para a quarta vizinhança – Cidade Fictícia.....	179
Figura 9-41: Significância do Indicador LISA para a quarta vizinhança – Cidade Fictícia ..	179
Figura 9-42: Mapa de <i>Clusters</i> LISA para a quarta vizinhança – Cidade Fictícia	180
Figura 9-43: Correlação do Indicador LISA para a quinta vizinhança – Cidade Fictícia.....	180
Figura 9-44: Significância do Indicador LISA para a quinta vizinhança – Cidade Fictícia ..	181
Figura 9-45: Mapa de <i>Clusters</i> LISA para a quinta vizinhança – Cidade Fictícia	181
Figura 9-46: Correlação do Indicador LISA para a sexta vizinhança – Cidade Fictícia	182
Figura 9-47: Significância do Indicador LISA para a sexta vizinhança – Cidade Fictícia....	182
Figura 9-48: Mapa de <i>Clusters</i> LISA para a sexta vizinhança – Cidade Fictícia	183
Figura 9-49: Correlação do Indicador LISA para a sétima vizinhança – Cidade Fictícia	183
Figura 9-50: Significância do Indicador LISA para a sétima vizinhança – Cidade Fictícia..	184
Figura 9-51: Mapa de <i>Clusters</i> LISA para a sétima vizinhança – Cidade Fictícia.....	184
Figura 9-52: Correlação do Indicador LISA para a oitava vizinhança – Cidade Fictícia.....	185
Figura 9-53: Significância do Indicador LISA para a oitava vizinhança – Cidade Fictícia ..	185
Figura 9-54: Mapa de <i>Clusters</i> LISA para a oitava vizinhança – Cidade Fictícia.....	186
Figura 9-55: Correlação do Indicador LISA para a nona vizinhança – Cidade Fictícia.....	186
Figura 9-56: Significância do Indicador LISA para a nona vizinhança – Cidade Fictícia	187
Figura 9-57: Mapa de <i>Clusters</i> LISA para a nona vizinhança – Cidade Fictícia.....	187
Figura 9-58: Correlação do Indicador LISA para a décima vizinhança – Cidade Fictícia	188
Figura 9-59: Significância do Indicador LISA para a décima vizinhança – Cidade Fictícia.	188
Figura 9-60: Mapa de <i>Clusters</i> LISA para a décima vizinhança – Cidade Fictícia.....	189
Figura 9-61: Correlação do Indicador LISA para a segunda vizinhança – Centro, São Paulo – São Paulo.....	189
Figura 9-62: Significância do Indicador LISA para a segunda vizinhança – Centro, São Paulo – São Paulo.....	190
Figura 9-63: Mapa de <i>Clusters</i> LISA para a segunda vizinhança – Centro, São Paulo – São Paulo.....	190
Figura 9-64: Correlação do Indicador LISA para a terceira vizinhança – Centro, São Paulo – São Paulo.....	191

Figura 9-65: Significância do Indicador LISA para a terceira vizinhança – Centro, São Paulo – São Paulo	191
Figura 9-66: Mapa de <i>Clusters</i> LISA para a terceira vizinhança – Centro, São Paulo – São Paulo	192
Figura 9-67: Correlação do Indicador LISA para a quarta vizinhança – Centro, São Paulo – São Paulo	192
Figura 9-68: Significância do Indicador LISA para a quarta vizinhança – Centro, São Paulo – São Paulo	193
Figura 9-69: Mapa de <i>Clusters</i> LISA para a quarta vizinhança – Centro, São Paulo – São Paulo	193
Figura 9-70: Correlação do Indicador LISA para a quinta vizinhança – Centro, São Paulo – São Paulo	194
Figura 9-71: Significância do Indicador LISA para a quinta vizinhança – Centro, São Paulo – São Paulo	194
Figura 9-72: Mapa de <i>Clusters</i> LISA para a quinta vizinhança – Centro, São Paulo – São Paulo	195
Figura 9-73: Correlação do Indicador LISA para a sexta vizinhança – Centro, São Paulo – São Paulo	195
Figura 9-74: Significância do Indicador LISA para a sexta vizinhança – Centro, São Paulo – São Paulo	196
Figura 9-75: Mapa de <i>Clusters</i> LISA para a sexta vizinhança – Centro, São Paulo – São Paulo	196
Figura 9-76: Correlação do Indicador LISA para a sétima vizinhança – Centro, São Paulo – São Paulo	197
Figura 9-77: Significância do Indicador LISA para a sétima vizinhança – Centro, São Paulo – São Paulo	197
Figura 9-78: Mapa de <i>Clusters</i> LISA para a sétima vizinhança – Centro, São Paulo – São Paulo	198

LISTA DE TABELAS

Quadro 2-1: Indicadores Globais de Autocorrelação Espacial.....	34
Quadro 2-2: Indicadores Locais de Autocorrelação Espacial.	36
Quadro 3-1: Expressões dos modelos de semivariograma teóricos transitivos mais comuns..	44
Tabela 4-1: Medidas descritivas do banco de dados – Cidade Fictícia.....	47
Tabela 4-2: Medidas descritivas do banco de dados - Centro, São Paulo – SP	49
Tabela 5-1: Parâmetros dos semivariogramas experimentais – Cidade Fictícia	64
Tabela 5-2: Parâmetros do modelo dos semivariogramas teóricos – Cidade Fictícia.....	64
Tabela 5-3: Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal – Cidade Fictícia.....	66
Tabela 5-4: Medidas descritivas das semivariâncias teórica para dados aleatórios – Cidade Fictícia	67
Tabela 5-5: Valores críticos por significância desejada para os valores de semivariância	68
Tabela 5-6: Semivariâncias padronizadas pelo patamar de semivariância dos dados aleatórios, o indicador proposto corresponde aos dados teóricos da direção principal – Cidade Fictícia .	69
Tabela 5-7: Teste de significância (95% de confiança) para os indicadores propostos – Cidade Fictícia	70
Tabela 5-8: Comparação de valores do indicador, z e status da hipótese nula (para confiança de 95%) para índice I de Moran e indicador proposto – Cidade Fictícia.....	70
Tabela 5-9: Parâmetros dos semivariogramas experimentais – Centro, São Paulo – SP	73
Tabela 5-10: Parâmetros do modelo dos semivariogramas teóricos – Centro, São Paulo – SP	74
Tabela 5-11: Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal – Centro, São Paulo – SP.....	75
Tabela 5-12: Medidas descritivas das semivariâncias teórica para dados aleatórios – Centro, São Paulo – SP.....	76
Tabela 5-13: Valores críticos, por significância desejada, para os valores de semivariância – Centro, São Paulo – SP.....	77
Tabela 5-14: Semivariâncias padronizadas pelo patamar de semivariância dos dados aleatórios, o indicador proposto corresponde aos dados teóricos da direção principal – Centro, São Paulo – SP.....	78

Tabela 5-15: Teste de significância (95% de confiança) para os indicadores propostos – Centro, São Paulo – SP	79
Tabela 5-16: Comparação de valores do indicador, z e status da hipótese nula (para significância de 95%) para índice I de Moran e indicador proposto – Centro, São Paulo – SP	79
Tabela 6-1: Medidas descritivas das semivariâncias teóricas para dados aleatórios de escolha modal – Cidade Fictícia	89
Tabela 6-2: Valores críticos por significância desejada para os valores de semivariância – Cidade Fictícia	90
Tabela 6-3: Medidas descritivas das semivariâncias teóricas para dados aleatórios de média de viagens por automóvel – Centro, SP – SP	104
Tabela 6-4: Valores críticos por significância desejada para os valores de semivariância – Centro, SP – SP.....	105
Quadro 6-5: Zonas (nome e numeração) identificadas pelo índice LISA e pelo indicador SIVAR-L para a primeira vizinhança, a direção relativa corresponde à direção do recorte..	112
Quadro 7-1: Comparações metodológicas entre índice de Moran e SIVAR	120
Tabela 9-1: Coeficientes do modelo de regressão linear múltipla – Cidade Fictícia.....	138
Tabela 9-2: Coeficientes do modelo de regressão linear múltipla – Centro, São Paulo – SP	138

SUMÁRIO

1. INTRODUÇÃO	25
1.1 Caracterização do Problema	25
1.2 Objetivos.....	27
1.3 Estrutura do trabalho	27
2. ANÁLISE ESPACIAL	29
2.1 Considerações Iniciais	29
2.2 Conceitos	32
2.2.1 Dependência espacial	32
2.2.2 Autocorrelação Espacial	32
2.2.3 Covariância.....	33
2.3 Tipos de Dados Espaciais	33
2.4 Indicadores de Autocorrelação Espacial.....	34
2.4.1 Indicadores Globais de Autocorrelação Espacial	34
2.4.2 Indicadores Locais de Autocorrelação Espacial.....	35
2.4.3 Visualização de Dependência Espacial	36
3. GEOESTATÍSTICA	39
3.1.1 Variável Regionalizada.....	40
3.1.2 Semivariograma.....	40
3.1.3 Obtenção do semivariograma experimental	41
3.1.4 Obtenção do semivariograma teórico	43
3.1.5 Krigagem	44
3.1.6 Aplicações de Geoestatística em estudos de Demanda por Transportes: as lacunas e motivações para realização do presente estudo	45
4. MATERIAIS E MÉTODO	47
4.1 Materiais	47
4.1.1 Bancos de Dados	47
4.1.1.1 Cidade Fictícia	47
4.1.1.2 Centro, São Paulo – SP	48
4.1.2 Aplicativos utilizados	50
4.2 Método.....	51

4.2.1	Indicador Global	52
4.2.1.1	Padronização de distâncias	54
4.2.1.2	Determinação da primeira vizinhança	54
4.2.1.3	Definição dos parâmetros para cálculo dos semivariogramas experimentais	55
4.2.1.4	Cálculo de semivariância para diversas vizinhanças	56
4.2.1.5	Modelagem do semivariograma teórico	56
4.2.1.6	Elaboração do Teste de Hipótese	57
4.2.1.6.1	Simulação	58
4.2.1.6.2	Distribuição Populacional	58
4.2.1.6.3	Pontos críticos	58
4.2.1.7	Padronização das semivariâncias	59
4.2.1.8	Teste de Significância para os indicadores	59
4.2.1.9	Comparação com índice I de Moran	60
4.2.2	Indicador Local	60
5.	RESULTADOS E DISCUSSÕES – INDICADOR GLOBAL	63
5.1	Cidade fictícia	63
5.1.1	Padronização de distâncias e determinação da primeira vizinhança	63
5.1.2	Definição dos parâmetros para cálculo dos semivariogramas experimentais	63
5.1.3	Cálculo de semivariância para diversas vizinhanças	64
5.1.4	Modelagem do semivariograma teórico	64
5.1.5	Elaboração do Teste de Hipótese	66
5.1.5.1	Simulação	66
5.1.5.2	Distribuição Populacional	66
5.1.5.3	Pontos críticos	67
5.1.6	Padronização das variâncias	68
5.1.7	Teste de Significância	69
5.1.8	Comparação com indicadores globais comumente utilizados na literatura	70
5.2	Centro, São Paulo – SP	73
5.2.1	Padronização de distâncias e determinação da primeira vizinhança	73
5.2.2	Definição dos parâmetros dos semivariogramas	73
5.2.3	Cálculo de semivariância para diversas vizinhanças	74
5.2.4	Modelagem do semivariograma teórico	74
5.2.5	Elaboração do Teste de Hipótese	76
5.2.5.1	Simulação	76

5.2.5.2	Distribuição Populacional.....	76
5.2.5.3	Pontos críticos.....	77
5.2.6	Padronização das semivariâncias.....	78
5.2.7	Teste de Significância.....	78
5.2.8	Comparação com indicadores globais conhecidos	79
5.3	Vantagens e desvantagens do SIVAR-G.....	80
6.	RESULTADOS E DISCUSSÕES – INDICADOR LOCAL	83
6.1	Cidade fictícia.....	83
6.1.1	Padronização de distâncias e determinação da primeira vizinhança	83
6.1.2	Definição dos parâmetros para cálculo dos semivariogramas experimentais	83
6.1.3	Cálculo de semivariância experimentais para diversas vizinhanças	83
6.1.4	Modelagem dos semivariogramas teóricos.....	85
6.1.5	Elaboração do Teste de Hipótese.....	88
6.1.5.1	Simulação.....	88
6.1.5.2	Distribuição Populacional.....	89
6.1.5.3	Pontos críticos.....	89
6.1.6	Padronização das variâncias teóricas.....	90
6.1.7	Teste de Significância.....	91
6.1.8	Comparação com indicadores locais comumente utilizados na literatura.....	92
6.2	Centro, São Paulo – SP.....	99
6.2.1	Padronização de distâncias e determinação da primeira vizinhança	99
6.2.2	Definição dos parâmetros para cálculo dos semivariogramas experimentais	99
6.2.3	Cálculo de semivariância experimentais para diversas vizinhanças	99
6.2.4	Modelagem dos semivariogramas teóricos.....	101
6.2.5	Elaboração do Teste de Hipótese.....	104
6.2.5.1	Simulação.....	104
6.2.5.2	Distribuição Populacional.....	104
6.2.5.3	Pontos críticos.....	105
6.2.6	Padronização das semivariâncias teóricas	106
6.2.7	Teste de Significância.....	108
6.2.8	Comparação com indicadores locais comumente utilizados na literatura.....	109
6.3	Vantagens e Desvantagens do SIVAR-L.....	116
7.	CONCLUSÕES E RECOMENDAÇÕES	119
7.1	Síntese do trabalho e conclusões	119

7.2	Recomendações.....	124
8.	REFERÊNCIAS BIBLIOGRÁFICAS	127
9.	LISTA DE ANEXOS E APÊNDICES.....	135
	Anexo A – Banco de dados Centro, São Paulo – SP	135
	Apêndice A – Regressão Linear.....	138
	Apêndice B – Semivariogramas experimentais e teóricos – Indicador Global.....	139
	Apêndice C – Exemplo de Cálculo de Indicador Global	141
	Apêndice D – Exemplo de Cálculo de Indicador Local.....	148
	Apêndice E – Mapas correspondentes aos semivariogramas teóricos padronizados SIVAR-L	162
	Apêndice F – Mapas correspondentes ao índice LISA	176

1. INTRODUÇÃO

1.1 Caracterização do Problema

Demanda por transportes corresponde à demanda gerada pelo desejo de locomover pessoas ou bens de um lugar a outro. Esse deslocamento ocorre no espaço urbano, sendo assim, a distribuição de atividades no espaço é essencial para compreender os fluxos de viagens existentes (ORTÚZAR; WILLUMSEN, 2011).

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; SILVA; CARVALHO, 2006).

Entre as escolhas realizadas pelos usuários, encontra-se a escolha do modo de transporte utilizado, o qual possui influência direta na eficiência geral da cidade (ORTÚZAR; WILLUMSEN, 2011). Por exemplo, o impacto no aumento de congestionamentos dependendo de maior escolha pela utilização de automóveis pelos usuários.

O processo de escolha do modo de transporte utilizado depende de características da viagem, do usuário e do sistema de transporte (BEN-AKIVA; LERMAN, 1985). Abordagens atuais enfatizam a necessidade de inclusão de componentes espaciais, pois se sabe 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; PENDYALA; BHAT, 2010).

Assim sendo, as viagens, realizadas diariamente, estão associadas às condições socioeconômicas dos usuários e à localização da residência e local de destino dos deslocamentos (RODRIGUES DA SILVA, 2008). Sabe-se que informações como renda do usuário, posse de automóvel, valor monetário da viagem e tempo de viagem apresentam estrutura espacial. Por consequência, variáveis que dependem dessas informações, como escolha modal, média de viagens por domicílio e média de viagens por modo automóvel por domicílio, por exemplo, também possuem dependência espacial.

A análise de dados de transporte urbano inclui métodos estatísticos descritivos, como métodos exploratórios e gráficos, mapeamento espacial, estatística espacial, análise de regressão, entre outros (BLACK; PÁEZ; SUTHANAYA, 2002).

O uso de modelos espaciais para estimativa de variáveis que possuam dependência espacial, mais especificamente – para o caso deste trabalho – variáveis relativas à demanda por transportes, requer uma investigação prévia e exploratória subjacente à dependência espacial destes dados.

Uma forma de avaliação da associação espacial de dados é através de indicadores, investigando se existe alguma tendência espacial das variáveis analisados. Confirmando-se essa influência, essa informação pode ser utilizada em modelos espaciais confirmatórios para previsão de viagens.

Indicadores consistem em um único número que comprova a existência de autocorrelação espacial e quantifica as relações de dependência espacial. A maioria dos índices diz respeito à similaridade entre valores da mesma variável em diferentes localizações e são aplicáveis somente a variáveis quantitativas, como é o caso do índice I de Moran global (1950), família de estatísticas G (GESTIS; ORD, 1992) e índice I de Moran local ou LISA (ANSELIN, 1995).

Dentre as técnicas de análise espacial, também se encontra a Geoestatística, a qual se baseia no conceito de variável regionalizada para analisar dependência espacial. Esse método vem sendo aplicado recentemente à demanda por transportes, inclusive a problemas relativos à escolha modal, como em Pitombo *et al.* (2015) e Lindner e Pitombo (2018).

Na Geoestatística, é medida a dissimilaridade dos valores da mesma variável em pontos de localizações diferentes, através do semivariograma. O semivariograma é um gráfico que relaciona semivariância, a distâncias fixas, entre os pares de pontos analisados. A técnica é aplicável a variáveis quantitativas e qualitativas.

Enquanto os indicadores espaciais apresentam o resultado na forma de um número, a geoestatística reflete o fenômeno de autocorrelação espacial através do gráfico semivariograma, que relaciona semivariância e distância entre pares de pontos.

Tanto os índices de análise espacial quanto o semivariograma consideram intervalos de valores de distâncias relevantes para os cálculos, excluindo os demais pontos que não pertencem a esses intervalos. No caso dos indicadores, a análise é feita para uma distância previamente especificada. A avaliação pode ser realizada para outras distâncias, porém essa não é uma exigência do método. Já através do semivariograma, é possível realizar a análise de todas as vizinhanças (distâncias fixas previamente especificadas) até a estacionariedade. Diversos estudos realizam análises espaciais utilizando indicadores e Geoestatística concomitantemente, considerando os métodos complementares (FORTIN; DRAPEAU; LEGENDRE, 1989; GOOVAERTS, 2010).

O presente trabalho consiste em uma análise exploratória de dados espaciais, produzindo indicadores baseados no semivariograma experimental e teórico. 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.

Os indicadores desenvolvidos correspondem à família de *Spatial Indicators based on Variographic Concepts*, ou SIVAR. Assim, o indicador global corresponde a SIVAR-G e o indicador local a SIVAR-L.

1.2 Objetivos

Propor um método para cálculo de indicadores globais e locais de associação espacial através do conceito de semivariograma;

Desenvolver um indicador de associação espacial apropriado para diferentes tipos de dados, diversas direções e capaz de identificar a estrutura espacial dos dados analisados;

Comparar os indicadores desenvolvidos a indicadores tradicionalmente conhecidos na literatura, com objetivo de validar o método proposto.

1.3 Estrutura do trabalho

Este trabalho está organizado em nove capítulos, incluindo essa introdução. O segundo capítulo trata da revisão de literatura a respeito da análise espacial e indicadores de associação espacial, apresentando seus conceitos, métodos de cálculo e representações. O terceiro capítulo aborda a geoestatística, explorando conceitos e trabalhos essenciais para a compreensão do método proposto. O quarto capítulo trata dos materiais, bancos de dados e aplicativos, e método, etapas realizadas para o desenvolvimento dos indicadores global e local baseados em conceitos variográficos.

O quinto capítulo apresenta os resultados obtidos através da aplicação do procedimento metodológico proposto para o indicador global de ambos os bancos de dados. Da mesma forma, o sexto capítulo apresenta os resultados correspondentes ao indicador local desenvolvido. O sétimo capítulo expõe as conclusões e recomendações.

O oitavo capítulo apresenta as referências bibliográficas utilizadas neste trabalho. Por fim, o nono capítulo trata dos anexos e apêndices, compreendendo resultados de cálculos intermediários, códigos utilizados e exemplos de aplicação.

2. ANÁLISE ESPACIAL

2.1 Considerações Iniciais

Compreender a distribuição espacial de fenômenos que ocorrem no espaço é uma questão relevante para diversas áreas do conhecimento, como saúde, ambiente, geologia, agronomia, entre outros. No campo de análise de demanda por transportes, muitas decisões realizadas pelos usuários ocorrem no espaço.

Os primeiros trabalhos relacionados à análise espacial foram frutos da revolução quantitativa da geografia, a qual ocorreu nas décadas de 1950 e 1960 (ANSELIN, 2010). Nesse período a geografia passou de um estudo regional para uma ciência espacial, desenvolvendo modelos preditivos de forma a compreender a estrutura e organização das relações espaciais (BERRY; MARBLE, 1968).

Até a metade da década de 1970, diversos autores abordaram a estimação de modelos espaciais. Tobler (1970) estabeleceu a “Primeira Lei da Geografia”, a qual considera que todos os objetos estão relacionados no espaço, sendo que objetos mais próximos entre si estão mais relacionados que objetos mais distantes.

Durante os anos que se seguiram, as pesquisas trataram principalmente sobre testes para resíduos de autocorrelação espacial, desenvolvimento de modelos espaciais, métodos básicos de estimação e alguns trabalhos iniciais em modelos de espaço-tempo (ANSELIN, 2010). Um dos produtos dessa época foi o índice I de Moran, concebido por Moran (1950), sendo a estatística desenvolvida apropriadamente anos depois por Cliff e Ord (1972).

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 (KELEJIAN; PRUCHA, 2001).

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 apresentado, em seguida, no capítulo 3). 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 o aplicado na estatística c de Geary (CLIFF; ORD, 1981).

No final da década de 1980, a falta de aplicativos adequados para analisar informações espaciais analiticamente era citado como maior entrave para a adoção de uma perspectiva espacial no trabalho empírico (ANSELIN; GRIFFITH, 1988). A partir da década de 1990, os avanços computacionais permitiram novas abordagens, inclusive a análise de estatísticas locais, permitindo a identificação de áreas de dependência espacial dentro do conjunto de dados.

Getis e Ord (1992) desenvolveram a família de estatísticas G , composta pela estatística global $G(d)$ e pelas estatísticas locais $G_i(d)$ e $G_i^*(d)$. No estudo, Getis e Ord (1992) recomendam que a técnica seja utilizada em conjunto com o índice I de Moran e as estatísticas locais são especialmente úteis na detecção de *outliers* e “bolsões” de dependência espacial.

Inúmeros trabalhos surgiram nessa época, sempre comparando as novas técnicas ao já consagrado índice I de Moran. Por exemplo: estatística de diagnóstico de autocorrelação localizada, intituladas LADS (NASS; GARFINKLE, 1991), testes de autocorrelação espacial em termos de perturbação em modelos de regressão (KELEJIAN; ROBINSON, 1992), testes de autocorrelação espacial utilizando multiplicadores de Lagrange (ANSELIN *et al.*, 1996).

Anselin (1995) desenvolveu uma nova classe de indicadores locais de associação espacial (LISA), sendo possível a sua utilização na decomposição do indicador global de Moran na contribuição de cada informação. O indicador desenvolvido busca evidenciar “bolsões” de não estacionaridade, similar às estatísticas G de Getis e Ord (1992) e avaliar a influência de localizações individuais na magnitude de uma estatística global, eventualmente identificando *outliers*.

O índice I de Moran e a estatística LISA de Anselin (1995) consagraram-se como as abordagens mais utilizadas para análise espacial global e local, respectivamente.

Dessa forma, historicamente a análise exploratória de variáveis quantitativas foi realizada através de diferentes métodos, com bases de cálculo diferentes. Todos os métodos utilizados até então pressupõem normalidade dos dados, não sendo adequados para aplicação a variáveis qualitativas. Além disso, os métodos consideram análises omnidirecionais (360°) ao redor dos pontos, retornando um valor correspondente à relação do ponto de análise e seus vizinhos (pontos a uma determinada distância do ponto de análise).

A análise exploratória de variáveis qualitativas é relativamente menos desenvolvida do que de variáveis contínuas. Essa preferência ocorreu devido à gama de técnicas para análise de resíduos de regressão, desenvolvidas ao longo dos anos. Os índices I de Moran e c de Geary são exemplos desse tipo de análise (RUIZ; LÓPEZ; PÁEZ, 2010).

Segundo Fortin, Dale e Hoef (2002), para o caso de espaçamento irregular em duas dimensões – caso correspondente aos dados obtidos através de pesquisas de análise de demanda por transportes – os métodos que podem ser aplicados a variáveis quantitativas são muitos. Como métodos de análise quantitativa são citados os índices comentados anteriormente, semivariância (geoestatística), SADIE, teste de Mantel multivariado, Análise de Superfície de Tendência, *Splines* e Krigagem, enquanto que para análise qualitativa cita-se apenas a técnica de Dimensão Fractal.

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 das ferramentas 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 do resultado. Na geoestatística, as semivariâncias são dadas segundo a distância entre pontos na forma do gráfico semivariograma, não sendo aplicado um teste de significância sobre o diagrama.

Fortin, Drapeau e Legendre (1989) investigaram a autocorrelação espacial aplicada à ecologia. Utilizaram diferentes métodos de análise espacial, como coeficientes de autocorrelação espacial (Moran e Geary) e geoestatística (semivariogramas e krigagem) para comparar amostras de diferentes tamanhos para estrutura de um conjunto de dados de densidade arbórea do bordo de açúcar em uma floresta no Québec. Concluiu-se que as técnicas podem ser aplicadas de forma complementar, de forma a compreender o fenômeno estudado através de diferentes perspectivas.

Diversos autores seguiram a mesma linha, realizando análise de dados espaciais através de índices e análise geoestatística. Zhang e McGrath (2003) analisaram a concentração de carbono em solos orgânicos. Machiwal *et al.* (2012) analisaram o nível de água em lençóis freáticos e o grau de exploração dos mesmos, utilizando a análise espacial como método complementar de planejamento e gestão de recursos. Nanos *et al.* (2001) utilizaram técnicas de análise espacial para estudar a variação de resinas em pinheiros. Goovaerts (2010) estudou dados agregados e desagregados de mortalidade por câncer. Toepfer *et al.* (2007) aplicaram as técnicas a estudos entomológicos de pragas de plantações.

Nesse trabalho propõe-se que os resultados dos semivariogramas, provenientes da geoestatística, possam ser interpretados na forma de indicadores. Para isso é essencial compreender os conceitos envolvidos nas ferramentas de análise espacial citadas.

2.2 Conceitos

2.2.1 Dependência espacial

Tobler (1970) estabelece como primeira lei da geografia: todos os objetos no espaço estão relacionados, mas objetos mais próximos são mais parecidos que objetos distantes entre si. Dependência espacial pode ser compreendida como 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. Os demais conceitos apresentados são uma decorrência da dependência espacial existente entre os elementos no espaço.

2.2.2 Autocorrelação Espacial

A dependência espacial pode ser mensurada através da autocorrelação espacial. O termo é derivado do conceito de correlação, a qual mensura a relação entre duas variáveis. Dessa forma, autocorrelação mede o grau que uma mesma variável aleatória, medida em pontos diferentes do espaço, tende a mudar (DRUCK *et al.*, 2004).

Para se calcular a autocorrelação espacial podem-se utilizar diferentes indicadores, sempre sendo uma estatística de produtos cruzados do tipo:

$$\Gamma(d) = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) \xi_{ij} \quad (2.1)$$

Em que: $\Gamma(d)$ autocorrelação dada uma distância (d); w_{ij} matriz de pesos segundo distância (d); ξ_{ij} matriz de correlação entre variáveis aleatórias.

O índice reflete a relação entre variáveis aleatórias z_i e z_j como produto de duas matrizes. Dada a distância (d), é possível elaborar uma matriz de pesos w_{ij} , a qual é dada segundo uma regra de contiguidade (válida apenas para análise de áreas), distância específica ou k-vizinhos mais próximos. A matriz ξ_{ij} depende do indicador utilizado.

Valores significativos de índices de autocorrelação espacial indicam que existe dependência espacial, o que inviabiliza a base teórica dos procedimentos de inferências estatística tradicionais, que exigem independência das amostras (DRUCK *et al.*, 2004).

2.2.3 Covariância

A covariância é a medida de interdependência entre duas variáveis aleatórias. Em análise espacial, a covariância, assim como a autocorrelação, mede a relação entre valores de uma mesma variável em pontos diferentes no espaço (YAMAMOTO; LANDIM, 2013).

A determinação da covariância pode depender de uma distância e direção pré-definidas. Em caso de diferentes variabilidades espaciais a partir de diferentes direções, tem-se um fenômeno ou variável anisotrópica. Caso não haja alteração de variabilidade espacial segundo a direção, o fenômeno é isotrópico, uniforme em todas as direções (YAMAMOTO; LANDIM, 2013).

2.3 Tipos de Dados Espaciais

Druck *et al.* (2004) caracterizam os dados utilizados usualmente em análise espacial em três tipos: eventos ou padrões pontuais, superfícies contínuas e áreas com contagens e taxas agregadas.

Eventos ou padrões pontuais são fenômenos representados por ocorrências, ou seja, observações pontuais. As análises mais comuns compreendem a determinação de padrões e agregação de dados. São exemplos: localização de acidentes, localização de crimes, ocorrência de doenças, localização de espécies animais e vegetais, entre outros.

Na análise de superfícies contínuas são estimadas superfícies baseadas em amostras de campo (pontuais), que podem ter distribuição regular ou irregular. Para sua interpretação são aplicadas técnicas de interpolação, medidas de incerteza e técnicas de geoestatística, como a krigagem, para determinação das superfícies.

Áreas com contagens e taxas agregadas compreendem dados de levantamentos sociodemográficos. Originalmente os dados referem-se a indivíduos (pontuais), porém, por razões de confidencialidade, os dados são agregados em áreas, onde se supõe que existe homogeneidade entre os indivíduos pertencentes a mesma área. Os dados deveriam ser agregados segundo a similaridade dos indivíduos, porém muitas vezes os critérios são operacionais não havendo garantias que os dados sejam homogêneos nas áreas delimitadas. São exemplos de unidades espaciais de análise de dados agregados: setores censitários, zonas de endereçamento postal, zonas de tráfego, municípios, bairros, entre outros.

2.4 Indicadores de Autocorrelação Espacial

Os indicadores representam a autocorrelação espacial, através de apenas um número. Nesse trabalho são apresentados os indicadores mais usuais na literatura, ou seja, o índice I de Moran (1950), índice *c* de Geary (1954), família de estatísticas G de Getis e Ord (1992), índice LISA para análise local (ANSELIN, 1995).

2.4.1 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 índice de Moran (1950), índice *c* de Geary (1954) e estatística global G de Getis e Ord (1992). As suas equações são apresentadas, em seguida, no Quadro 2-1.

Quadro 2-1: Indicadores Globais de Autocorrelação Espacial.

Estatística	Equação	
Índice I de Moran	$I(d) = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$	(2.2)
Estatística <i>c</i> de Geary	$c = \frac{n-1}{2 \sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)} \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)(x_i - x_j)^2}{\sum_{i=0}^n (x_i - x_i)^2}$	(2.3)
Estatística G(d)	$G(d) = \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij}(d)x_i x_j}{\sum_{i=0}^n \sum_{j=0}^n x_i x_j}$	(2.4)

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.

É essencial 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 mede a correlação espacial através de uma medida de covariância na forma de produto cruzado. O índice varia de -1 a +1, valores negativos indicam autocorrelação espacial negativa (os valores dos vizinhos são diferentes entre si, como em um tabuleiro de xadrez), valores próximos de zero indicam ausência de autocorrelação espacial e valores positivos indicam autocorrelação espacial positiva (os valores dos vizinhos são semelhantes entre si).

A hipótese nula do teste é a não existência de autocorrelação espacial, resultando em valor nulo e correspondendo a um padrão aleatório (KREMPI, 2004).

Camargo (1998) destaca os métodos de validação estatística utilizados como o teste de pseudo-significância, no qual são geradas permutações dos valores de atributo associados às regiões, cada permutação gerando um novo arranjo espacial, os valores são redistribuídos entre as áreas. Como apenas um dos arranjos corresponde à situação observada, é construída uma situação empírica do índice I , caso o valor de I corresponda ao extremo da distribuição simulada é dito que o evento é estatisticamente significativo.

A estatística c de Geary (1954) utiliza o quadrado das diferenças entre os pares de valores, de forma similar ao cálculo do variograma experimental. O índice varia de 0 a 2, sendo que 0 a 1 corresponde a autocorrelação espacial positiva e 1 a 2 autocorrelação espacial negativa.

Getis e Ord (1992) desenvolveram família de estatísticas G , a qual pertence à estatística global $G(d)$. A estatística define um conjunto de vizinhos dentro de um raio pré-definido para serem utilizados efetivamente nos cálculos. Baseando-se nesses vizinhos, são identificados “hot spots” (variável com valor alto vizinhos com valor alto) e “cool spots” (variável com valor baixo e vizinhos com valor baixo), tal como a representação gráfica do índice I de Moran. Representa um índice de concentração espacial, pois identifica apenas autocorrelação espacial positiva.

2.4.2 Indicadores Locais de Autocorrelação Espacial

Quando o volume de dados é grande, é comum que os indicadores globais de autocorrelação espacial não identifiquem bolsões de variação espacial ou *outliers*. Indicadores locais permitem que estes fenômenos sejam estudados em detalhe. Os indicadores locais produzem valores específicos para cada área ou ponto, permitindo a identificação e agrupamentos (CAMARGO, 1998). As suas equações são apresentadas, em seguida, no Quadro 2-2.

A estatística LISA possibilita a decomposição do indicador global índice I de Moran, permitindo compreender a contribuição de localizações individuais para o cálculo do índice global I de Moran.

A família de estatísticas G de Getis e Ord (1992) possui dois índices de análise local, G_i e G_i^* . Para cada localidade a estatística G_i mede se o ponto em questão é rodeado por valores altos, *hot spots*, ou valores baixos *cool spots*. Apenas os vizinhos de i são considerados no

cálculo. Caso queira-se calcular o índice com a inclusão da região i o índice é modificado para G_i^* .

Quadro 2-2: Indicadores Locais de Autocorrelação Espacial.

Estatística	Equação	
LISA	$I_i = \frac{x_i \sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n x_j^2}$	(2.5)
$G_i(d)$	$G_i(d) = \frac{\sum_{j=0}^n w_{ij}(d) x_j}{\sum_{j=0}^n x_j}, \quad \text{para } j \neq i$	(2.6)
$G_i^*(d)$	$G_i^*(d) = \frac{\sum_{j=0}^n w_{ij}(d) x_j}{\sum_{j=0}^n x_j}, \quad \forall j$	(2.7)

Em que: x_i valor da variável normalizada na região i ; w_{ij} são os elementos da matriz de pesos para a ordem de vizinhança ou lag d ; x_j elementos considerados como vizinhança e utilizados para o cálculo.

2.4.3 Visualização de Dependência Espacial

Uma das técnicas de visualização de dependência espacial é o diagrama de espalhamento de Moran, o qual compara os valores normalizados de uma variável com a média dos seus vizinhos. O diagrama consiste em um gráfico bidimensional de valores normalizados versus média dos vizinhos.

O índice I é interpretado como um coeficiente de regressão linear, relacionando a inclinação da reta de regressão de WZ (variável Z ponderada por pesos da matriz de vizinhança) em Z . Assim, quanto mais próximo da reta estabelecida (quadrantes Q1 e Q2) entende-se que a associação espacial entre os dados é positiva, enquanto que dados em quadrantes aos quais a reta não pertence (quadrantes Q3 e Q4) indica associação espacial negativa.

O quadrante Q1 corresponde a pontos com valores acima da média geral e com a média dos vizinhos desses pontos acima da média geral. O quadrante Q2 corresponde ao inverso, pontos com valores abaixo da média geral e com a média dos vizinhos desses pontos abaixo da média geral. Dessa forma, a região analisada e os vizinhos possuem valores semelhantes e sugerem associação espacial positiva.

O quadrante Q3 corresponde a pontos com valores acima da média geral e com a média dos vizinhos desses pontos abaixo da média geral. O quadrante Q4 corresponde ao inverso, pontos com valores abaixo da média geral e com a média dos vizinhos desses pontos acima da média geral. Dessa forma, a região analisada e os vizinhos possuem valores distintos e sugerem associação espacial negativa.

O indicador local de autocorrelação espacial LISA também pode ser expresso na forma de mapa, no qual indica quais regiões apresentam correlação local significativamente diferente

do restante dos dados. Nesse caso, o mapa é dado em graus de significância, cabendo ao analista reconhecer quais informações são relevantes.

Pode-se construir o mapa de espalhamento de Moran e de LISA, da mesma forma que os dados estão divididos em quadrantes, as áreas são identificadas como “Alto-Alto”, “Baixo-Baixo”, “Alto-Baixo” e “Baixo-Alto”, correspondendo a Q1, Q2, Q3 e Q4, conforme ilustrado na Figura 2-1.

Figura 2-1: Diagrama de Espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, censo de 1991 (CAMARGO, 1998).

3. GEOESTATÍSTICA

O engenheiro sul africano Daniel G. Krige (1951) desenvolveu os princípios da geoestatística para estimar, de uma forma mais adequada, as propriedades das minas nas quais trabalhava. Para ele, informações baseadas somente na variância não seriam suficientes para caracterizar as amostras. Era necessário considerar a distância entre os pontos coletados (CAMARGO, 1998).

Baseado nos estudos de Krige (1951), Matheron (1970) desenvolveu a teoria das variáveis regionalizadas, dando forma à geoestatística. A geoestatística reconhece a variabilidade espacial em grande e pequena escala, ou seja, ela modela ambas tendências espaciais e correlação espacial (CRESSIE, 1991).

A partir da década de 1980, a geoestatística teve uma larga aplicação nos campos de mineração, geologia, agricultura, cartografia, climatologia, engenharia florestal, análise espacial de crimes, epidemiologia, geotécnica e pedologia (YAMAMOTO; LANDIM, 2013).

Na Estatística Clássica, as variáveis são consideradas aleatórias, seus valores são considerados independentes entre si e parâmetros como média e desvio padrão são suficientes para descrever as características principais das amostras.

Caso as variáveis estejam distribuídas em um ambiente de duas ou três dimensões estas podem apresentar uma estrutura de dependência espacial, a qual não é contemplada pela Estatística Clássica. 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).

Na Geoestatística, cada variável representa uma distribuição de probabilidades e ocorrência de valores (YAMAMOTO; LANDIM, 2013). A caracterização das variáveis é realizada através de semivariogramas, ferramenta que considera as particularidades espaciais da mesma.

Um fenômeno que apresenta estrutura espacial é chamado de fenômeno regionalizado, da mesma forma, variáveis que dizem respeito a esse fenômeno são chamadas de variáveis regionalizadas. As próximas seções trazem uma síntese de conceitos básicos da geoestatística. Descrições mais detalhadas podem ser verificadas em Matheron (1970), Journel e Huijbregts (1978), Journel (1989), Soares (2000) e Wackernagel (2003).

3.1.1 Variável Regionalizada

Matheron (1970) desenvolveu a teoria das variáveis regionalizadas, as quais são funções numéricas com dependência espacial e que não podem ser explicadas através de funções determinísticas. Seu objetivo é não só caracterizar e auxiliar na compreensão da correlação espacial, mas também fornecer fundamentos para produção de uma superfície de estimativas (YAMAMOTO; LANDIM, 2013).

A variação espacial de uma variável regionalizada pode ser expressa como a soma de três elementos: uma componente estrutural (valor médio ou tendência constante); uma componente aleatória espacialmente correlacionada; e um ruído aleatório ou erro residual (BURROUGH, 1986). A função que representa uma variável regionalizada é dada por:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'' \quad (3.1)$$

Em que: $Z(x)$ variável regionalizada dependente de x ; $m(x)$ função determinística que descreve a componente estrutural da função Z em x ; $\varepsilon'(x)$ termo estocástico que varia localmente e depende espacialmente de x ; ε'' ruído aleatório não correlacionado.

Atualmente, é possível aplicar técnicas de geoestatística a variáveis quantitativas (contínuas e discretas) e qualitativas (nominais, ordinais e binárias).

A estrutura espacial do fenômeno regionalizado pode apresentar tendência segundo uma direção, configurando uma estrutura anisotrópica. Caso não haja essa tendência o fenômeno é dito isotrópico (YAMAMOTO; LANDIM, 2013). Para analisar em detalhes a estrutura espacial de uma variável regionalizada utiliza-se a ferramenta do semivariograma.

3.1.2 Semivariograma

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 e se estabilize em um patamar, a partir do qual não há mais dependência espacial.

O termo utilizado pode ser variograma ou semivariograma. Matheron (1970) pretendia calcular as diferenças entre um ponto e seu vizinho determinado por uma distância h dada por:

$$[Z(x + h) - Z(x)] \quad (3.2)$$

Porém o valor desejado era metade desse, o prefixo semi corresponde à divisão por dois, como demonstrado a seguir.

$$2\gamma(h) = E[Z(x + h) - Z(x)]^2 \quad (3.3)$$

$$\gamma(h) = \frac{1}{2} E[Z(x + h) - Z(x)]^2 = \frac{1}{2n} \sum_{i=1}^n [Z(x + h) - Z(x)]^2 \quad (3.4)$$

Em que: $2\gamma(h)$ função variograma em h ; $\gamma(h)$ função semivariograma em h ; h distância entre pares de pontos; $Z(x)$ variável aleatória Z em x ; n número de elementos contabilizados.

Portanto, $2\gamma(h)$ é chamado de variograma e $\gamma(h)$ semivariograma. Na literatura é possível encontrar o semivariograma chamado de variograma, porém sempre utilizando a divisão por dois.

3.1.3 Obtenção do semivariograma experimental

Para a cálculo do semivariograma experimental, 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 ponto utilizados para o cálculo incluem o ponto de origem do vetor e os pontos dentro da área delimitada pelas tolerâncias, conforme apresentado na Figura 3-1.

Figura 3-1: Esquema de cálculo de semivariogramas experimentais para distribuições irregulares. Adaptado de Camargo (1998).

Para cada etapa do processo, é escolhido um ponto como base, a partir dele são geradas áreas segundo a direção fixada e a distância h , a qual é repetida a análise para as distâncias $2h$, $3h$, e assim por diante. Em cada análise é aplicada a função semivariograma para cada par de pontos, constituído por ponto encontrado na área delimitada e o ponto base.

Todos os pontos são analisados para diferentes distâncias, múltiplas de h , de forma a ser possível construir um gráfico relacionando semivariância γ e distância entre os pares (h).

Baseado nas hipóteses de média constante e estacionaridade, espera-se que as observações mais próximas tenham um comportamento mais semelhante entre si (DRUCK *et al.*, 2004). Os valores da função semivariograma devem aumentar segundo a distância até alcançar o ponto onde não é possível observar dependência espacial (patamar).

Druck *et al.* (2004) e Journel e Huijbregts (1978) descrevem os pontos notáveis dos semivariogramas que permitem a sua interpretação a seguir:

- Alcance ou amplitude (a): distância total observada em que as amostras apresentam correlação espacial.
- Patamar (C): valor da semivariância que corresponde à amplitude, da mesma forma, até alcançar esse valor as amostras apresentam correlação espacial. A partir do alcance o valor da função semivariograma pode variar em torno da semivariância máxima, chamada patamar.
- Efeito Pepita (C_0): em uma situação ideal, quanto menor a distância h , menor deve ser a função semivariograma até o valor nulo, quando h for igual a zero. Porém, devido a descontinuidades do semivariograma para pequenas distâncias, pode ocorrer que, para pequenos valores de h , a função semivariograma possua valores maiores do que o esperado. A causa do efeito pepita não é completamente conhecida, podendo ter origem em erros de medição, variabilidades de pequena escala ou simplesmente falta de informação.
- Contribuição (C_1): corresponde à diferença entre o patamar (C) e o efeito pepita (C_0).

Pode acontecer do efeito pepita ser tão dominante que a função semivariograma não aumenta de valor com o aumento da distância. Nesse caso as informações possuem tantos ruídos aleatórios (ϵ'') que nenhuma interpretação e interpolação é possível, o semivariograma pode ser representado pelo valor do efeito pepita puro (BURROUGH, 1986).

Após gerar os pontos experimentais do semivariograma é desejável obter um modelo matemático que melhor os represente através do ajuste do semivariograma experimental para um teórico (Figura 3-2).

Figura 3-2: Esquema semivariograma teórico. Adaptado de Burrough (1986)

3.1.4 Obtenção do semivariograma teórico

O ajuste dos modelos é realizado de forma manual ou automática (através do método dos mínimos quadrados), analisando quais são os melhores modelos que se adaptam a cada trecho do conjunto de dados considerando as suposições apresentadas no item anterior. Os modelos teóricos usualmente aplicados são divididos em com e sem patamar.

Modelos com patamar são denominados transitivos. Os modelos mais usuais para representá-los são efeito pepita puro, esférico, exponencial e gaussiano, apresentados no Quadro 3-1.

Nos modelos sem patamar o valor da função semivariograma continua aumentando com o valor da distância h. Esses modelos são utilizados para fenômenos com capacidade infinita de dispersão, utiliza-se a função potência para representá-los com a calibração dos parâmetros α e β (DRUCK *et al.*, 2004).

$$\gamma(h) = \alpha h^\beta, \text{ com } 0 < \beta < 2 \quad (3.5)$$

Para alguns fenômenos, é necessário utilizar uma composição de modelos aninhados para que a representação seja aquela mais fiel possível. Pode-se combinar modelos de tipos diferentes ou até do mesmo tipo com parâmetros diferentes.

Quadro 3-1: Expressões dos modelos de semivariograma teóricos transitivos mais comuns.

Tipo	Amplitude	Equação	
Efeito Pepita Puro	0	$\gamma(h) = C$	(3.6)
Esférico	a	$\gamma(h) = \begin{cases} C_0 + \frac{C}{2} \left[3 \left(\frac{h}{a} \right) - \left(\frac{h}{a} \right)^3 \right] & \text{para } h < a \\ C_0 + C & \text{para } h \geq a \end{cases}$	(3.7)
Exponencial	3a	$\gamma(h) = C_0 + C \left[1 - \exp \left(-\frac{h}{a} \right) \right]$	(3.8)
Gaussiano	$a\sqrt{3}$	$\gamma(h) = C_0 + C \left[1 - \exp \left(-\left(\frac{h}{a} \right)^2 \right) \right]$	(3.9)

Em que: $\gamma(h)$ função semivariograma em h; C_0 efeito pepita; C patamar; h distância entre pares de pontos; a alcance.

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. Para obter essa informação são calculados semivariogramas para diversos ângulos e comparados entre si. Caso seja identificada anisotropia, é necessário investigar as direções principais dos padrões espaciais.

A anisotropia geométrica ocorre quando é definido um conjunto de semivariogramas teóricos com patamares de amplitudes diferentes dependendo da direção analisada. Na anisotropia zonal o alcance é o mesmo, porém os patamares obtidos são diferentes. A anisotropia mista é a combinação dos fenômenos descritos anteriormente (YAMAMOTO; LANDIM, 2013).

3.1.5 Krigagem

A Krigagem corresponde ao objetivo último da Geoestatística, estimar atributos em locais sem informações. Consiste em um processo de estimação linear como uma função dos valores de variáveis próximas. Em vez de uma matriz de pesos, utiliza-se os parâmetros obtidos através de um semivariograma teórico.

No presente trabalho a krigagem não será abordada, pois o objetivo é limitado à interpretação de semivariogramas experimentais e teóricos.

3.1.6 Aplicações de Geoestatística em estudos de Demanda por Transportes: as lacunas e motivações para realização do presente estudo

A aplicação de técnicas de geoestatística à Demanda por Transportes é uma linha de pesquisa relativamente nova. São encontrados trabalhos tanto para estimativa de geração de viagens (ROCHA; LINDNER; PITOMBO, 2017; LINDNER *et al.*, 2016; ROCHA; PITOMBO; SALGUEIRO, 2016; PITOMBO; SOUSA, 2009) quanto para escolha modal (LINDNER; PITOMBO, 2018; PITOMBO *et al.*, 2015; LINDNER, 2015; GOMES *et al.*, 2016).

Recentemente, desenvolveram-se métodos sequenciais para estimação de escolha modal aplicando algoritmos de Árvore de Decisão e Krigagem Ordinária (COSTA, 2013; PITOMBO *et al.*, 2015; PITOMBO; DA COSTA; SALGUEIRO, 2015). Ainda no trabalho de Pitombo *et al.* (2015), a Árvore de Decisão gerada foi validada com a calibração de um modelo logit multinomial (MNL).

Lindner e Pitombo (2018) apresentaram uma interpolação espacial de probabilidades de escolha modal baseada na abordagem clássica logit multinomial (MNL) e abordagem espacial com aplicação geoestatística para estimar dados em coordenadas geográficas de valores não pesquisados.

Lindner (2015) e Gomes *et al.* (2016) utilizaram Krigagem Indicativa para determinação da superfície de probabilidades de escolha modal. Lindner (2015) modelou semivariogramas experimentais e teóricos a partir da variável binária “Escolha Modal” para obtenção da superfície de probabilidades por Krigagem Indicativa. A mesma autora comparou os resultados obtidos a partir da Krigagem Indicativa às estimativas realizadas através de Regressão Logística Binomial. Apesar da segunda técnica apresentar maior taxas de acertos, através da Krigagem Indicativa, além de ser possível estimativa em pontos previamente não amostrados, é necessário um volume de informação muito menor como *input* (coordenadas geográficas e escolha modal).

Já Gomes *et al.* (2016) comparam os resultados obtidos através da Krigagem Indicativa com os resultados obtidos através de algoritmo de Árvore de Decisão e Krigagem Ordinária (método sequencial), analisando vantagens e desvantagens de cada ferramenta. Os trabalhos aqui citados pertencem ao mesmo grupo de pesquisa (Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo) e têm o objetivo comum de estimar probabilidades de escolhas modais ou viagens produzidas e atraídas, com o auxílio de modelagem geoestatística.

O presente trabalho, no entanto, preenche uma lacuna importante quando comparado aos estudos anteriores, pois propõe um procedimento metodológico para cálculo e teste de significância de indicadores de associação espacial, baseados nos conceitos da geoestatística. A estimativa de uma variável regionalizada, através de estatística espacial (incluindo a geoestatística), só é possível se houver dependência espacial da variável em estudo. Assim, este trabalho aborda a análise prévia à estimativa, que seria a análise exploratória da dependência espacial da variável. Neste projeto, os conceitos do semivariograma auxiliam na composição do indicador de associação espacial.

4. MATERIAIS E MÉTODO

4.1 Materiais

4.1.1 Bancos de Dados

Os bancos de dados utilizados no presente projeto compreendem dados desagregados por indivíduos e agregados a partir de uma malha determinada. São eles: uma cidade fictícia, construída de modo a apresentar uma estrutura espacial bem definida (ASSIRATI *et al.*, 2016), e o Centro da cidade de São Paulo – SP, dados da Pesquisa de Mobilidade 2012, executada pela Companhia do Metropolitano de São Paulo (METRÔ, 2012). Este último (um banco de dados real) foi utilizado com o propósito de validação metodológica.

4.1.1.1 Cidade Fictícia

A cidade fictícia é proveniente do trabalho de Assirati *et al.* (2016). O banco de dados consiste numa população sintética, com valores de variáveis espacialmente correlacionados. Tais dados foram obtidos através do procedimento de interpolação bi-linear.

A variável de análise corresponde à escolha modal, a qual é qualitativa binária. Caso o modo escolhido seja ônibus, a variável corresponde ao valor zero. Caso o modo seja automóvel, a variável corresponde ao valor um. A distribuição geográfica desta variável é apresentada na Figura 4-1.

O banco de dados utilizado possui 580 pontos com as medidas descritivas de seus atributos quantitativos representados na Tabela 4-1.

Tabela 4-1: Medidas descritivas do banco de dados – Cidade Fictícia

	X	Y	Distância (Km)	Tempo Ônibus (Min.)	Tempo Carro (Min.)	Preço Ônibus (Reais)	Preço Carro (Reais)	Renda (Escala 1 a 10)
Média	5,20	2,68	2,47	5,40	3,70	1,99	2,10	5,43
Desvio Padrão	2,93	1,45	1,31	2,58	1,97	0,70	1,11	1,47
Valor Máximo	10,00	5,00	5,52	11,12	8,29	3,00	4,70	10,00
Valor Mínimo	0,00	0,15	0,05	1,08	0,07	1,50	0,04	1,00

Fonte: (ASSIRATI *et al.*, 2016)

Figura 4-1: Representação da variável Escolha Modal na cidade fictícia, produzida por Assirati *et al.* (2016)

4.1.1.2 Centro, São Paulo – SP

Como exemplo de um banco de dados real utilizou-se a variável média de viagens por modo automóvel, por domicílios, dentro de uma determinada área. O método descrito a seguir faz parte da pesquisa de doutorado, ainda em curso da aluna Samille Rocha, no Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos – USP.

Com a intenção de analisar uma região real, com continuidade espacial, realizou-se um recorte da cidade de São Paulo. As zonas de tráfego provenientes da Pesquisa de Mobilidade (METRÔ, 2012) foram agrupadas em *clusters* com valores de renda similares. Assim, obteve-se uma região central de valores, com provável dependência espacial, utilizada para a análise.

Os domicílios foram agregados em grades de 500 por 500 m, apresentadas na Figura 4-2. Para cada uma dessas grades calculou-se a média da variável “viagens por automóvel” por domicílios contidos na célula. Os valores agregados (Figura 4-2) foram associados ao centroide de sua grade, como apresentado na Figura 4-3. Para este trabalho, os dados utilizados correspondem aos apresentados na forma de centroides.

O banco de dados analisado possui 905 observações e as seguintes medidas descritivas, apresentadas na Tabela 4-2.

Tabela 4-2: Medidas descritivas do banco de dados - Centro, São Paulo – SP

	X	Y	Média de Viagens por Automóvel (un.)	Média de Viagens Total (un.)	Média de Viagens por Automóvel/Média de Viagens Total
Média	328890,065	7389807,754	2,285	5,484	0,356
Desvio Padrão	3938,011	4752,444	1,906	2,732	0,269
Valor Máximo	337039,847	7399146,066	10	17	1
Valor Mínimo	320910,716	7379511,328	0	0	0

Figura 4-2: Número médio de viagens por modo automóvel por quadrícula de 500 por 500 m.

Figura 4-3: Número médio de viagens por modo automóvel, associado aos centroides das quadriculas de 500 por 500 m.

A associação do presente banco de dados às zonas de tráfego da pesquisa de Mobilidade de 2012 pode ser observada no Anexo A – Banco de dados Centro, São Paulo – SP.

4.1.2 Aplicativos utilizados

Os aplicativos utilizados nesse trabalho encontram-se descritos a seguir.

- *QGIS*: aplicativo gratuito e *open-source* de Sistemas de Informações Geográficas (GIS). Foi utilizado para visualização, análise das variáveis do banco de dados e produção de mapas.
- *Microsoft Excel*: *software* de processamento de banco de dados. Foi utilizado para análise dos dados e suporte para rotinas em outros aplicativos.
- *R Studio*: *software* de desenvolvimento integrado produzido pelo *R Core Team* (2017), priorizando a programação para gráficos e cálculos estatísticos. Foi utilizado para: cálculos de parâmetro de suavização da Regressão Geograficamente Ponderada, através da biblioteca *spgwr* (BIVAND; YU, 2017); semivariogramas, através da biblioteca *geoR* (RIBEIRO; DIGGLE, 2016); cálculo do indicador Moran Global através da

biblioteca *lctools* (KALOGIROU, 2017); ajuste dos modelos do semivariograma teórico local *minpack.lm* (ELZHOV *et al.*, 2016); cálculo do indicador Moran Local através da biblioteca *ncf* (BJORNSTAD, 2016); além da biblioteca gráfica *ggplot2* (WICKHAM, 2009), e de bibliotecas de configuração de dados como geográficos *sp* (BIVAND; PEBESMA; GOMEZ-RUBIO, 2013) e *lattice* (SARKAR, 2008).

- *SGeMS – Stanford Geoestatistical Modeling Software*: aplicativo *open-source* utilizado para análises geoestatísticas que possibilitou a modelagem complexa do semivariograma global.

4.2 Método

Nesse trabalho são desenvolvidos dois indicadores SIVAR (*Spatial Indicators based on Variographic Concepts*). O primeiro corresponde ao indicador global, SIVAR-G, que remete a estrutura espacial do banco de dados como um todo, sendo apresentado na forma da variância teórica padronizada global. O segundo corresponde ao indicador local, SIVAR-L, o qual apresenta uma análise da estrutura espacial de cada elemento da base de dados, sendo apresentado na forma de semivariância teórica padronizada unitária.

Ambos os indicadores seguem as premissas descritas a seguir.

- Para cada vizinhança de distância pré-definida é obtido um valor de semivariância entre o ponto e seus vizinhos;
- Este é avaliado, através de teste de hipótese, a fim de verificar se existe autocorrelação espacial nos dados analisados.

O indicador L_i , indicado na Equação 4.1, é dado em função dos valores de uma determinada variável na localização i e dos valores dos vizinhos a i para essa variável. A vizinhança é determinada através do parâmetro de suavização utilizado na técnica de Regressão Geograficamente Ponderada.

$$L_i = f(y_i, y_{j_i}) \quad (4.1)$$

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 do indicador ser superior ao valor crítico seja inferior ao nível de significância utilizado.

$$\text{Prob}[L_i > \delta_i] \leq \alpha_i \quad (4.2)$$

Em que: L_i indicador; δ_i valor crítico; α_i significância escolhida ou nível de pseudo-significância, por exemplo, resultado de um teste aleatório.

Como este trabalho possui um caráter exploratório, optou-se por comparar os indicadores obtidos com indicadores tradicionais globais e locais, como Moran e LISA. Observa-se que os indicadores citados são aplicáveis com confiabilidade apenas para dados quantitativos. Contudo, considerou-se necessária a comparação do indicador SIVAR com outros indicadores consolidados, de forma a verificar o desempenho de análise dos dados. Dessa forma, Moran e LISA foram aplicados a dados binários, o que não é adequado, já que pressupõem que a variável analisada possui distribuição normal.

Destaca-se que os índices Moran e LISA se baseiam em um cálculo de covariância, enquanto o SIVAR baseia-se em variância. Dessa forma, esperam-se resultados condizentes entre si, porém não iguais, já que os cálculos realizados são diferentes.

4.2.1 Indicador Global

Através do indicador global de associação espacial proposto, SIVAR-G, é possível realizar a análise de conjunto de dados para diversas vizinhanças.

A Figura 4-4 apresenta o procedimento metodológico para obtenção do SIVAR-G na forma de um fluxograma. As etapas que compreendem a obtenção do indicador são descritas nas subseções subsequentes.

Figura 4-4: Fluxograma do método para o Indicador Global de Associação Espacial proposto

4.2.1.1 Padronização de distâncias

Para tornar possível a comparação entre variáveis diferentes, optou-se por padronizar as distâncias entre pontos. A padronização permite que a diferença de escala entre bancos de dados não interfira na interpretação dos dados.

4.2.1.2 Determinação da primeira vizinhança

Visando determinar a distância que compreende os pontos mais relevantes para a análise espacial e buscando automatizar a escolha do lag entre pontos, utilizado para o cálculo do semivariograma, foi utilizado o conceito de parâmetro de suavização (*bandwidth*), proveniente da técnica de Regressão Geograficamente Ponderada (RGP).

A técnica de RGP 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 através da determinação do parâmetro de suavização (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002; FISCHER; GETIS, 2009).

O parâmetro de suavização controla a variância da função de ponderação (*kernel*) utilizada para determinar a matriz de pesos de cada ponto utilizada na determinação dos demais parâmetros do modelo. As funções de ponderação podem ser fixas ou variáveis (FISCHER; GETIS, 2009).

As funções fixas determinam o parâmetro de suavização através de uma distância fixa válida para todos os pontos analisados. As funções variáveis determinam k pontos vizinhos que compreendem a área de influência dos pontos. Desse modo, a distância considerada varia para cada ponto (FISCHER; GETIS, 2009). Como o objetivo é obter a distância da vizinhança ideal apenas as funções de ponderação fixas foram consideradas.

Caso o parâmetro de suavização possua um valor alto ele irá produzir parâmetros muito similares ao longo da área de estudo, caso o valor seja muito baixo ele irá produzir parâmetros com tamanha variação local que será difícil distinguir onde existem padrões de fato. Dessa forma, é necessário determinar critérios de seleção do parâmetro de suavização, de forma a encontrar um valor médio adequado entre os extremos. Os critérios de seleção mais populares são as minimizações através da validação cruzada (CV) e do critério de informação Akaike

(AIC), os quais geralmente não apresentam diferenças significativas entre si (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

Na técnica de validação cruzada o parâmetro de suavização é dado pelo valor que minimiza a equação a seguir (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002):

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{i \neq i})^2 \quad (4.3)$$

Em que: n número de pontos analisados; y_i valor observado da variável no ponto i; $\hat{y}_{i \neq i}$ valor ajustado de y_i com a observação para a localização i omitida do processo de estimação.

Na técnica AIC, o valor do parâmetro de suavização é o que minimiza a seguinte equação (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002):

$$AIC = -2 \log L(\hat{\theta}|y) + 2k \quad (4.4)$$

Em que: $L(\hat{\theta}|y)$ função de verossimilhança maximizada para o modelo; k número de parâmetros estimados para o modelo.

Para a determinação da primeira vizinhança, o parâmetro de suavização deve ser calculado através das técnicas de validação cruzada (CV) e do critério de informação Akaike (AIC), adotando-se o menor valor entre eles.

Se mais que 10% das observações não possuem vizinhos o valor utilizado de *bandwidth* deve ser descartado, já que não abrange adequadamente os dados analisados.

Caso o primeiro valor escolhido, o menor entre CV e AIC, não atenda esse critério, deve-se utilizar o maior valor entre os dois. Se, ainda assim, o critério não for atingido, deve-se adotar o menor valor de *bandwidth* calculado e multiplicá-lo por 2, 3, 4, etc. até que o critério seja alcançado.

4.2.1.3 Definição dos parâmetros para cálculo dos semivariogramas experimentais

A análise variográfica da variável inicia-se através da definição dos parâmetros para cálculo dos semivariogramas experimentais. São definidos os parâmetros apresentados na Figura 3-1, ou seja, lag (distância entre pontos), tolerância de lag (neste trabalho foi adotado um valor equivalente à metade do lag utilizado), tolerância angular, largura de banda (geralmente metade do lag utilizado) e direções nas quais serão calculadas os semivariogramas

experimentais (principal e ortogonal). Vale ressaltar que, com os valores escolhidos para lag e tolerância de lag, são determinados indicadores para faixas ou intervalos de distâncias.

A distância entre pontos utilizada (lag) foi previamente determinada através do parâmetro de suavização proveniente da técnica de RGP, descrito anteriormente.

As direções de cálculo podem variar, mais comum é realizar medições em intervalos de 30° ou 45°, de forma a ter uma visão geral do comportamento espacial dos dados. Neste trabalho, foram utilizadas as direções 0°, 30°, 60°, 90°, 120° e 150° e omnidirecional.

4.2.1.4 Cálculo de semivariância para diversas vizinhanças

Fixados os parâmetros na etapa anterior, os semivariogramas experimentais são calculados através dos aplicativos *R* e *SGeMS*.

Dados anisotrópicos possuem uma direção de máxima continuidade, ou seja, a direção na qual o patamar estabiliza-se para o menor valor de semivariância possível. Essa é identificada como direção principal, a direção imediatamente a 90° desta é identificada como direção ortogonal. Para o cálculo do semivariograma experimental, os resultados de um vetor negativo são os mesmos de um vetor positivo, desde que se mantenha a direção. Portanto, a direção ortogonal pode ser +90° ou -90° a partir da direção principal.

Apenas os dados da direção principal e ortogonal são utilizados para os cálculos posteriores.

4.2.1.5 Modelagem do semivariograma teórico

Os resultados do semivariograma experimental representam a associação espacial dos dados de um modo descontínuo. Dessa forma, faz-se necessário representar essa associação espacial na forma do semivariograma ajustado a um modelo teórico.

O semivariograma teórico deve ser ajustado para as direções principal e ortogonal, utilizando o mesmo efeito pepita, número de estruturas e tipo de função. Se a estrutura espacial for isotrópica a função semivariograma é a mesma para ambas as direções. Caso contrário existe anisotropia, as funções semivariograma da direção principal e da direção ortogonal apresentam alcances diferentes.

Utiliza-se uma função elipsoidal $\alpha(\theta)$, na qual o eixo maior corresponde ao alcance na direção principal e o eixo menor corresponde ao alcance na direção ortogonal. O valor de θ corresponde ao ângulo da direção principal. A Figura 4-5 apresenta um exemplo de elipsoide

de anisotropia, para a função semivariograma da direção principal utiliza-se o valor de $a1$ como alcance, enquanto que para a função semivariograma da direção ortogonal utiliza-se o valor de $a2$ como alcance.

Figura 4-5: Alcances do elipsoide de anisotropia

4.2.1.6 Elaboração do Teste de Hipótese

Assim como os demais indicadores apresentados, é necessário aplicar um teste de hipótese para descartar a aleatoriedade dos dados. O teste de hipótese utilizado corresponde a um teste de pseudo-significância, assim como aplicado para os indicadores de Moran (1950). Para estimar a significância do indicador proposto é necessário associar este a uma distribuição de probabilidades conhecida.

As hipóteses básicas do teste são formuladas como:

Hipótese Nula: A variável é espacialmente aleatória.

Hipótese Alternativa: A variável é espacialmente dependente. Existe um padrão espacial.

O objetivo desta etapa metodológica foi gerar uma distribuição de valores do indicador global proposto para o caso de aleatoriedade espacial. Essa distribuição é essencial para a determinação do teste de hipótese para avaliar a significância estatística do indicador SIVAR-G.

É realizada uma simulação do comportamento de semivariogramas teóricos calculados para dados que apresentam aleatoriedade. A distribuição da semivariância teórica é comparada à normal padrão e são determinados os pontos críticos que definem a aceitação ou rejeição da hipótese nula.

4.2.1.6.1 Simulação

Gerou-se um banco de dados com as mesmas coordenadas de longitude e latitude que o banco de dados original, porém a variável foi substituída por valores uniformemente aleatórios entre os valores mínimos e máximos da variável original. Em seguida, foi calculado e ajustado um semivariograma experimental omnidirecional e um semivariograma teórico para os dados simulados.

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

Esse procedimento foi repetido 100 vezes, gerando valores uniformemente aleatórios para cada iteração.

Os valores aleatórios e os semivariogramas foram produzidos utilizando o *software R* (R CORE TEAM, 2017).

4.2.1.6.2 Distribuição Populacional

Os valores de semivariância teórica para as 100 repetições foram analisados conjuntamente para cada uma das vizinhanças analisadas.

Em seguida, verificou-se se a distribuição de valores correspondia a uma distribuição normal através do teste de *Kolmogorov-Smirnov*, realizado através do *software R*. Sabe-se, pelo teorema do limite central, que para quão maior o número de observações, a tendência da média das mesmas converge para uma distribuição normal.

4.2.1.6.3 Pontos críticos

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 unilateral 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.

$$z = \frac{x - \mu}{\sigma} \quad (4.5)$$

Em que: z valor padronizado; x valor correspondente ao padronizado na distribuição original; μ média da distribuição original; σ desvio padrão da distribuição original.

Dessa forma, se o valor do indicador obtido estiver acima dos limites estabelecidos, a hipótese nula deve ser retida com um dado nível de significância. Caso o valor esteja abaixo desses limites, a hipótese nula é rejeitada e assume-se que há um padrão espacial.

4.2.1.7 Padronização das semivariâncias

De forma a tornar comparáveis semivariogramas de diferentes dados com diferentes patamares é necessário padronizar as semivariâncias. Essa padronização deve ser realizada de forma que o valor máximo possível para o patamar de qualquer semivariograma seja equivalente ao patamar típico de dados aleatórios.

O valor médio da semivariância corresponde ao patamar dos dados aleatórios e deve ser obtido através da análise dos dados realizada no item anterior.

O indicador proposto corresponde aos valores de variância teórica padronizada para a direção principal. Adotaram-se os dados teóricos, pois se observou que, por esses serem representados por um modelo, apresentam um comportamento mais uniforme, sem as mudanças abruptas de valores observadas nos dados do semivariograma experimental. Adotou-se a direção principal como o indicador, pois ela representa a direção de maior continuidade espacial dos dados, ou seja, de menor diferença entre valores, dando uma orientação clara da dependência espacial.

4.2.1.8 Teste de Significância para os indicadores

Através do teste de significância, é possível dizer se a hipótese nula de aleatoriedade é retida ou rejeitada.

Tendo calculado os valores dos pontos críticos e o valor dos indicadores devidamente padronizados, realiza-se um teste de significância, avaliando se para cada valor de indicador por vizinhança retém-se ou não a hipótese nula de aleatoriedade espacial.

4.2.1.9 Comparação com índice I de Moran

Finalmente, calcula-se o índice I de Moran para o conjunto de dados analisados. O valor obtido é comparado aos valores do indicador global proposto de forma a validar o procedimento metodológico proposto e compreender melhor os resultados obtidos.

4.2.2 Indicador Local

Através do indicador local de associação espacial proposto, SIVAR-L, é possível realizar a análise de cada um dos pontos do conjunto de dados, verificando seu comportamento para diversas vizinhanças. O procedimento metodológico para cálculo e obtenção do indicador local é similar ao apresentado para o indicador global, dessa forma as etapas do método são comentados de maneira sucinta a seguir.

Os procedimentos de padronização de distâncias e determinação da primeira vizinhança são análogos aos apresentados anteriormente.

A respeito da definição dos parâmetros para cálculo dos semivariogramas experimentais, pode-se realizar procedimento similar ao do indicador global, definindo-se diversas direções de análise e interpretando-as em conjunto, de forma a verificar as tendências à anisotropia para cada ponto. Porém, dependendo da direção e tolerâncias escolhidas, podem-se estabelecer regiões com poucos pares de pontos, comprometendo o cálculo e confiabilidade dos indicadores obtidos.

Nesse trabalho, optou-se por realizar uma análise omnidirecional (360°) para o indicador local, de forma a abranger um maior número de pontos na análise, evitando-se a ausência de vizinhos para certas configurações.

Definidos os parâmetros, é possível calcular as semivariâncias para as vizinhanças definidas em uma análise ponto a ponto. Os semivariogramas experimentais foram calculados através do *software R*.

Os semivariogramas experimentais foram ajustados aos modelos teóricos através do aplicativo *R*, utilizando a função *nlsLM*, mínimos quadrados não lineares. Tendo essas funções modeladas, ponto a ponto, pode-se avaliar a qualidade dos modelos através da correlação entre os dados observados e estimados.

O teste de hipótese desenvolvido é análogo ao apresentado para o indicador global. As diferenças correspondem à forma como é calculada a semivariância para a análise local e o número de resultados, já que se obtém uma distribuição para cada item do banco de dados.

Essas distribuições são analisadas em conjunto, possibilitando calcular média e desvio padrão e, conseqüentemente, os pontos críticos correspondentes a uma distribuição normal padrão.

Cada ponto terá seu semivariograma experimental e teórico característicos. Para possibilitar a comparação entre pontos, padronizam-se as semivariâncias teóricas dos pontos. O procedimento de determinação do patamar é similar ao do indicador global.

As demais etapas de determinação do indicador local através dos valores de semivariância teórica padronizada, teste de significância dos indicadores e comparação com outros índices seguem o mesmo procedimento descrito para o indicador global. Porém, cabe ressaltar que a análise é realizada sempre pontualmente e que o indicador tradicional utilizado, para fins comparativos, corresponde ao Moran Local, ou LISA.

5. RESULTADOS E DISCUSSÕES – INDICADOR GLOBAL

Nas próximas seções deste capítulo, são apresentados os resultados e discussões relativos ao indicador espacial global (SIVAR-G) para os dois bancos de dados em questão: (1) Variável binária relativa à escolha modal de uma cidade fictícia espacialmente estruturada, e (2) Média de viagens domiciliares por automóvel por célula 500 x 500m para região central de São Paulo (Pesquisa de Mobilidade de São Paulo, 2012).

5.1 Cidade fictícia

5.1.1 Padronização de distâncias e determinação da primeira vizinhança

As distâncias foram padronizadas, resultando em um banco de dados onde os valores de latitude e longitude variavam de 0 a 1. Em seguida, para determinação da primeira vizinhança relativa ao indicador, e com a finalidade de se obter o parâmetro de suavização, foi utilizado o critério de seleção de validação cruzada (CV) e critério de informação de Akaike (AIC) e a função de ponderação gaussiana fixa. Para a aplicação das funções, foi necessário supor um modelo de regressão linear entre as variáveis do banco de dados. Descartando as variáveis não significativas para o nível de confiança de 95%, obteve-se um modelo de previsão da variável probabilidade de escolha modal partindo dos dados de preço de modo ônibus, tempo de viagem no modo ônibus, renda e uma constante. O coeficiente de determinação do modelo corresponde a 0,719. Os parâmetros do modelo completo podem ser verificados no Apêndice A – Regressão Linear.

Através do *software R*, obteve-se o parâmetro de suavização ideal através do critério de seleção de validação cruzada igual a 0,038 e através do critério AIC igual a 0,032. Adotou-se o menor valor entre os dois (0,032) e verificou-se se mais de 90% das observações possuíam vizinhos, adotando esse valor como distância entre vizinhos.

5.1.2 Definição dos parâmetros para cálculo dos semivariogramas experimentais

A análise variográfica da variável escolha modal foi realizada para as direções 0°, 30°, 60°, 90°, 120° e 150° e omnidirecional. A Tabela 5-1, apresenta os parâmetros adotados para

cálculo dos semivariogramas experimentais para a cidade fictícia. A distância máxima de análise corresponde a 0,550, a partir da qual não é verificada mais associação espacial.

Tabela 5-1: Parâmetros dos semivariogramas experimentais – Cidade Fictícia

Semivariograma Experimental	
Lag	0,032
Tolerância Lag	0,016
Tolerância Ângular	22°
Largura de Banda	0,016

5.1.3 Cálculo de semivariância para diversas vizinhanças

Utilizando os parâmetros definidos no item anterior, calcularam-se os semivariogramas experimentais. Os resultados obtidos para cada uma das direções, comentadas no item anterior podem ser consultados no Apêndice B – Semivariogramas experimentais .

Os dados analisados são anisotrópicos, a direção principal corresponde a 0° e a direção ortogonal a 90°. Na seção seguinte são apresentados os variogramas experimentais e teóricos para estas direções.

Utilizou-se o *software SGeMS* devido a maior riqueza de detalhes na análise, porém o mesmo procedimento pode ser realizado diretamente no *software R* sem grande perda de qualidade.

5.1.4 Modelagem do semivariograma teórico

Como os dados analisados são anisotrópicos, o semivariograma teórico foi ajustado para as direções principal e ortogonal. A Tabela 5-2 apresenta os parâmetros do semivariograma teórico, o valor de alcance $a1$ corresponde à direção principal e o valor de alcance $a2$ à direção ortogonal.

Tabela 5-2: Parâmetros do modelo dos semivariogramas teóricos – Cidade Fictícia

Semivariograma Teórico	
Modelo	Esférico
Efeito Pepita (Co)	0,010
Contribuição (C)	0,450
Alcance $a1$	3,600
Alcance $a2$	0,500
θ	0°

Dessa forma, para a direção principal a equação utilizada corresponde a Equação 5.1 e para a direção ortogonal a equação corresponde a Equação 5.2.

$$\gamma(h) = \begin{cases} 0,010 + \frac{0,450}{2} \left[3 \left(\frac{h}{3,600} \right) - \left(\frac{h}{3,600} \right)^3 \right] & \text{para } h < 3,600 \\ 0,460 & \text{para } h \geq 3,600 \end{cases} \quad (5.1)$$

$$\gamma(h) = \begin{cases} 0,010 + \frac{0,450}{2} \left[3 \left(\frac{h}{0,500} \right) - \left(\frac{h}{0,500} \right)^3 \right] & \text{para } h < 0,500 \\ 0,460 & \text{para } h \geq 0,500 \end{cases} \quad (5.2)$$

A Figura 5-1 ilustra os semivariogramas experimentais e ajustados.

Figura 5-1: Semivariogramas experimentais e teóricos para a direção principal e ortogonal – Cidade Fictícia

A Tabela 5-3 sintetiza a análise variográfica e apresenta os valores obtidos de semivariância experimental e teórica para as dez primeiras vizinhanças, as quais possuem maior influência sobre o resultado de seus vizinhos, segundo as direções principal e ortogonal. O patamar obtido para esse caso corresponde à semivariância de 0,460.

Tabela 5-3: Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal – Cidade Fictícia

Vizinhança	Distância (h)	Semivariância (γ)			
		Experimental		Teórico	
		Direção Principal (0°)	Direção Ortogonal (90°)	Direção Principal (0°)	Direção Ortogonal (90°)
1	0,032	0,010	0,038	0,016	0,053
2	0,064	0,015	0,064	0,022	0,096
3	0,096	0,031	0,095	0,028	0,138
4	0,128	0,031	0,149	0,034	0,179
5	0,160	0,052	0,227	0,040	0,219
6	0,192	0,067	0,285	0,046	0,256
7	0,224	0,056	0,342	0,052	0,292
8	0,256	0,084	0,331	0,058	0,325
9	0,288	0,088	0,354	0,064	0,356
10	0,320	0,077	0,392	0,070	0,383

5.1.5 Elaboração do Teste de Hipótese

5.1.5.1 Simulação

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 “escolha modal” correspondem a números aleatórios zero ou um. O código utilizado na simulação é apresentado no Apêndice C – Exemplo de Cálculo de Indicador Global.

Para os dados aleatórios, os semivariogramas desenvolvidos são omnidirecionais, de forma a considerar quaisquer tendências espaciais que os novos dados possam apresentar. Os dados aleatórios não possuem estrutura espacial, portanto produziram semivariogramas com tendência constante, bem representados pelo modelo efeito pepita puro. Esse procedimento foi repetido 100 vezes, gerando novos dados aleatórios a cada repetição.

5.1.5.2 Distribuição Populacional

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.

Como não há correlação espacial, os resultados para todas as vizinhanças são similares, pois parte-se do pressuposto de que a variância entre os valores das variáveis em pontos diferentes não varia significativamente. A Tabela 5-4 apresenta os valores de média, desvio padrão, máximo e mínimo da semivariância calculada nas simulações para todas as vizinhanças. Optou-se por utilizar apenas os dados da semivariância correspondente ao modelo teórico, ou seja, os modelos efeito pepita puro.

Tabela 5-4: Medidas descritivas das semivariâncias teórica para dados aleatórios – Cidade Fictícia

Medidas descritivas	
Média	0,250
Desvio Padrão	1,461e-3
Valor Máximo	0,252
Valor Mínimo	0,244

Em seguida, aplicou-se o teste de normalidade *Kolmogorov-Smirnov* ao novo conjunto de dados, os valores obtidos correspondem a D de 0,143 e p-valor de 0,002. Considerando um nível de confiança de 95% rejeita-se a hipótese nula de normalidade. É provável que a hipótese de normalidade não tenha sido retida devido ao tipo de variável (binária). No entanto, optou-se pela suposição de normalidade da distribuição, especialmente considerando o alto número de observações, de forma a dar prosseguimento à obtenção do indicador proposto.

5.1.5.3 Pontos críticos

Se a distribuição encontrada corresponde à distribuição normal, é possível encontrar os valores críticos correspondentes para o nível de significância unilateral desejado. Os pontos críticos são obtidos através do cálculo que relaciona a distribuição produzida à distribuição normal padrão.

A Tabela 5-5 apresenta os valores críticos, por significância desejada, tendo como base os valores de semivariância obtidos através da análise de dados aleatórios. A Figura 5-2 apresenta o histograma dos dados, probabilidade acumulada e uma indicação do valor correspondente ao intervalo de confiança de 95%.

Tabela 5-5: Valores críticos por significância desejada para os valores de semivariância

Intervalo de confiança	z	p-valor (α)	Limite	Limite (dados padronizados)
80%	-0,845	0,20	0,249	0,995
85%	-1,035	0,15	0,248	0,994
90%	-1,290	0,10	0,248	0,992
95%	-1,645	0,05	0,248	0,990
99%	-2,330	0,01	0,247	0,986

Figura 5-2: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Cidade Fictícia

5.1.6 Padronização das variâncias

É possível observar, na Tabela 5-4, que o valor médio da variância dos dados aleatórios corresponde a 0,250. Dessa forma, esse valor corresponde ao novo valor máximo que o patamar do semivariograma dos dados analisados pode alcançar.

Assim, as novas equações para a direção principal e ortogonal são apresentadas a seguir, através da Equação 5.3 e Equação 5.4.

$$\gamma(h) = \begin{cases} 0,022 + \frac{0,978}{2} \left[3 \left(\frac{h}{3,600} \right) - \left(\frac{h}{3,600} \right)^3 \right] & \text{para } h < 3,600 \\ 1 & \text{para } h \geq 3,600 \end{cases} \quad (5.3)$$

$$\gamma(h) = \begin{cases} 0,022 + \frac{0,978}{2} \left[3 \left(\frac{h}{0,500} \right) - \left(\frac{h}{0,500} \right)^3 \right] & \text{para } h < 0,500 \\ 1 & \text{para } h \geq 0,500 \end{cases} \quad (5.4)$$

Realizando-se a padronização, é possível obter os valores apresentados na Tabela 5-6, sendo que os valores de variância teórica padronizada para a direção principal correspondem aos valores relativos ao indicador proposto SIVAR-G. Os limites correspondentes aos dados padronizados são apresentados na última coluna da Tabela 5-5, apresentada no item anterior.

A ideia da padronização das variâncias seria homogeneizar os valores obtidos do indicador, independentemente da escala da variável em análise. Observa-se, neste caso, que o zero corresponde à forte dependência espacial, enquanto o valor unitário corresponderia a ausência de dependência espacial, já que o SIVAR-G corresponde a um indicador de dissimilaridades entre valores da mesma variável em diferentes vizinhanças. Quanto mais próximo do zero, menores as dissimilaridades.

Tabela 5-6: Semivariâncias padronizadas pelo patamar de semivariância dos dados aleatórios, o indicador proposto corresponde aos dados teóricos da direção principal – Cidade Fictícia

Vizinhança	Distância (h)	Semivariância (γ)			
		Experimental		Teórico	
		Direção Principal (0°)	Direção Ortogonal (90°)	Direção Principal (0°)	Direção Ortogonal (90°)
1	0,032	0,021	0,082	0,035	0,116
2	0,064	0,032	0,140	0,048	0,209
3	0,096	0,067	0,207	0,061	0,300
4	0,128	0,068	0,324	0,074	0,389
5	0,160	0,114	0,494	0,087	0,475
6	0,192	0,145	0,620	0,100	0,558
7	0,224	0,122	0,743	0,113	0,635
8	0,256	0,183	0,719	0,126	0,707
9	0,288	0,192	0,769	0,139	0,773
10	0,320	0,167	0,852	0,152	0,833

5.1.7 Teste de Significância

Utilizando a Tabela 5-5 de valores críticos por significância desejada, é possível analisar os dados obtidos na Tabela 5-6, que apresenta o indicador proposto para diferentes vizinhanças. Dessa forma, os valores do indicador SIVAR-G, com os respectivos testes de significância para a hipótese nula de aleatoriedade espacial, são apresentados na Tabela 5-7.

Tabela 5-7: Teste de significância (95% de confiança) para os indicadores propostos – Cidade Fictícia

Vizinhança	Distância (h)	Direção Principal (0°)	H0
1	0,032	0,035	Rejeita
2	0,064	0,048	Rejeita
3	0,096	0,061	Rejeita
4	0,128	0,074	Rejeita
5	0,160	0,087	Rejeita
6	0,192	0,100	Rejeita
7	0,224	0,113	Rejeita
8	0,256	0,126	Rejeita
9	0,288	0,139	Rejeita
10	0,320	0,152	Rejeita

Observa-se que as dez primeiras vizinhanças tiveram a hipótese de aleatoriedade espacial rejeitada para confiança de 95%.

5.1.8 Comparação com indicadores globais comumente utilizados na literatura

O cálculo do índice de Moran pressupõe normalidade dos dados. Apesar da distribuição dos dados analisados ser binomial, para o cálculo do índice de Moran supõe-se que a distribuição seja equivalente à gaussiana devido ao grande número de observações (580 pontos).

A Tabela 5-8, a seguir, apresenta a relação entre o índice proposto e o índice de Moran, de forma a validar os resultados encontrados.

Tabela 5-8: Comparação de valores do indicador, z e status da hipótese nula (para confiança de 95%) para índice I de Moran e indicador proposto – Cidade Fictícia

Viz,	Dist,	Moran				SIVAR-G			
		Índice	Z	p-valor	H0	Índice	Z	p-valor	H0
1	0,032	0,902	36,316	0,000	Rejeita	0,035	-165,083	0,000	Rejeita
2	0,064	0,850	46,467	0,000	Rejeita	0,048	-162,852	0,000	Rejeita
3	0,096	0,846	53,372	0,000	Rejeita	0,061	-160,623	0,000	Rejeita
4	0,128	0,782	65,871	0,000	Rejeita	0,074	-158,394	0,000	Rejeita
5	0,160	0,699	74,867	0,000	Rejeita	0,087	-156,167	0,000	Rejeita
6	0,192	0,643	82,704	0,000	Rejeita	0,100	-153,941	0,000	Rejeita
7	0,224	0,583	85,043	0,000	Rejeita	0,113	-151,718	0,000	Rejeita
8	0,256	0,508	83,971	0,000	Rejeita	0,126	-149,497	0,000	Rejeita
9	0,288	0,434	81,824	0,000	Rejeita	0,139	-147,279	0,000	Rejeita
10	0,320	0,360	76,023	0,000	Rejeita	0,152	-145,064	0,000	Rejeita

O valor do índice I de Moran para os dados de escolha modal, para a cidade fictícia, corresponde a aproximadamente 0,902 para a primeira vizinhança. Como esperado, reflete alta autocorrelação espacial entre os dados. Os valores do índice decrescem à medida que a distância considerada aumenta, todos eles são rejeitados para 95% de confiança no teste de hipótese de Moran, ou seja, rejeita-se a hipótese nula de aleatoriedade e considera-se que há autocorrelação espacial entre os dados.

O indicador proposto possui um comportamento similar, utilizando os limites obtidos na Tabela 5-5, para 95% de confiança, rejeita-se a hipótese da variável ser espacialmente aleatória nas dez vizinhanças analisadas, assim como o índice I de Moran.

Na Figura 5-3 os índices são apresentados comparativamente por vizinhança. É possível observar que os índices são de natureza inversa ou complementar, pois o índice de Moran é baseado na similaridade dos dados, enquanto o índice proposto utiliza conceitos variográficos, baseando-se na dissimilaridade entre os dados.

Este banco de dados apresenta forte continuidade na direção principal 0° , com um comportamento anisotrópico, como pode ser observado comparando os resultados do indicador SIVAR-G (direção principal) com os resultados da semivariância padronizada da direção ortogonal na Figura 5-3. Observa-se que os valores correspondentes à direção ortogonal possuem um comportamento mais similar ao do Índice de Moran. Observa-se, então, que o indicador SIVAR-G é capaz de captar a forte continuidade dos dados na direção principal (0°), retornando valores que indicam uma associação espacial superior ao índice de Moran para cada vizinhança.

Figura 5-3: Comparação entre: a) os índices de Moran, indicador SIVAR-G e exposição do limite crítico para o nível de confiança de 95%, b) além dos valores já apresentados são representados os valores de semivariância para a direção ortogonal, de forma a comparar os resultados na direção principal e ortogonal – Cidade Fictícia

5.2 Centro, São Paulo – SP

5.2.1 Padronização de distâncias e determinação da primeira vizinhança

A exemplo do banco de dados anterior, as distâncias foram padronizadas e a primeira vizinhança determinada através dos mesmos critérios descritos anteriormente (critério de seleção de validação cruzada (CV) e critério de informação de Akaike (AIC) e a função de ponderação gaussiana fixa.

Foram obtidos valores similares (0,022) tanto para o critério de seleção de validação cruzada quanto para o critério AIC. No entanto, verificou-se que mais de 10% das observações não possuíam vizinhos para ambos os critérios.

Portanto, optou-se por utilizar o menor valor de *bandwidth* (0,022 por AIC) multiplicado por 2 (0,044), o qual satisfaz o critério de mais de 90% das observações possuírem vizinhos.

5.2.2 Definição dos parâmetros dos semivariogramas

Para o banco de dados da região central de São Paulo, também foi realizada a investigação da estrutura espacial da variável segundo as mesmas direções citadas anteriormente (0°, 30°, 60°, 90°, 120° e 150°) e omnidirecional. A distância máxima de análise corresponde a cerca de 0,800, a partir da qual não é observada mais dependência espacial. A Tabela 5-9 apresenta os parâmetros dos variogramas experimentais segundo as direções pesquisadas.

Tabela 5-9: Parâmetros dos semivariogramas experimentais – Centro, São Paulo – SP

Semivariograma Experimental	
Lag	0,044
Tolerância Lag	0,022
Tolerância Ângular	22°
Largura de Banda	0,022

5.2.3 Cálculo de semivariância para diversas vizinhanças

Observando os semivariogramas experimentais, obtém-se que a direção principal corresponde a 120° e a direção ortogonal a 30°. Os variogramas experimentais para as seis direções podem ser observados no Apêndice B – Semivariogramas experimentais .

5.2.4 Modelagem do semivariograma teórico

Os dados analisados são anisotrópicos, portanto o semivariograma teórico foi ajustado para as direções principal e ortogonal. A Tabela 5-10 apresenta os parâmetros do semivariograma teórico, o valor de alcance $a1$ corresponde à direção principal e o valor de alcance $a2$ a direção ortogonal.

Tabela 5-10: Parâmetros do modelo dos semivariogramas teóricos – Centro, São Paulo – SP

Semivariograma Teórico	
Modelo	Exponencial
Efeito Pepita (Co)	2,000
Contribuição (C)	1,800
Alcance $a1$	0,200
Alcance $a2$	0,100
θ	120°

Dessa forma, para a direção principal a equação utilizada corresponde a Equação 5.5 e para a direção ortogonal a equação corresponde a Equação 5.6.

$$\gamma(h) = 2,000 + 1,800 \left[1 - \exp \left(-\frac{h}{0,200} \right) \right] \quad (5.5)$$

$$\gamma(h) = 2,000 + 1,800 \left[1 - \exp \left(-\frac{h}{0,100} \right) \right] \quad (5.6)$$

A Figura 5-4 apresenta os semivariogramas experimentais e ajustados para ambas as direções.

Figura 5-4: Semivariogramas experimentais e teóricos para a direção principal e ortogonal – Centro, São Paulo – SP.

A Tabela 5-11 sintetiza a análise variográfica e apresenta os valores obtidos de semivariância experimental e teórica para as primeiras vizinhanças, as quais possuem maior influência sobre o resultado de seus vizinhos, segundo as direções principal e ortogonal. O patamar obtido para esse caso corresponde à semivariância de 3,800.

Tabela 5-11: Resumo dos valores de semivariância experimental e teórica para direção principal e ortogonal – Centro, São Paulo – SP

Vizinha nça	Distância (h)	Semivariância (γ)			
		Experimental		Teórico	
		Direção Principal (120°)	Direção Ortogonal (30°)	Direção Principal (120°)	Direção Ortogonal (30°)
1	0,044	2,462	2,771	2,357	2,515
2	0,088	2,608	3,082	2,643	2,974
3	0,133	2,860	3,236	2,873	3,269
4	0,177	3,198	3,266	3,057	3,459
5	0,221	3,128	3,574	3,204	3,581
6	0,265	3,258	3,695	3,322	3,659
7	0,310	3,378	3,688	3,417	3,709
8	0,354	3,477	3,653	3,493	3,742
9	0,398	3,590	3,753	3,554	3,763
10	0,442	3,720	3,800	3,603	3,776

5.2.5 Elaboração do Teste de Hipótese

5.2.5.1 Simulação

Como descrito para o exemplo anterior, foi gerado um banco de dados com as mesmas coordenadas que o banco de dados original região central de São Paulo. Os valores adotados para a variável média de viagens por automóvel correspondem a números aleatórios entre zero e dez, os valores mínimo e máximo do banco de dados original. O código utilizado na simulação é apresentado no Apêndice C – Exemplo de Cálculo de Indicador Global.

Os semivariogramas desenvolvidos são omnidirecionais, considerando tendências espaciais que os dados possam apresentar. Os dados aleatórios não apresentaram estrutura espacial, produzindo semivariogramas com tendência constante representados pelo modelo efeito pepita puro. Esse procedimento foi repetido 100 vezes, gerando novos dados aleatórios a cada repetição.

5.2.5.2 Distribuição Populacional

Optou-se por utilizar a distribuição dos dados de semivariância teórica, ou seja, os modelos efeito pepita puro. A Tabela 5-12 apresenta os valores da média, desvio padrão, máximo e mínimo das semivariâncias relativas aos modelos teóricos das 100 simulações.

Tabela 5-12: Medidas descritivas das semivariâncias teórica para dados aleatórios – Centro, São Paulo – SP

Medidas descritivas	
Média	8,292
Desvio Padrão	0,262
Valor Máximo	8,919
Valor Mínimo	7,415

Aplicou-se o teste de normalidade *Kolmogorov-Smirnov* ao novo conjunto de dados, os valores obtidos correspondem a D de 0,065 e p-valor de 0,797. Considerando um nível de confiança de 95% retém-se a hipótese nula, ou seja, os dados possuem uma distribuição similar à normal.

5.2.5.3 Pontos críticos

Assim como descrito anteriormente (subseção 4.2.1.6.3), é possível obter os valores correspondentes aos pontos críticos através da comparação da distribuição analisada à distribuição normal padrão. Se o valor do indicador obtido estiver acima dos limites estabelecidos, a hipótese nula deve ser retida com um dado nível de significância, ou seja, os dados não possuem associação espacial. Caso contrário, se H_0 for rejeitada, há autocorrelação espacial.

A Tabela 5-13 apresenta os valores críticos, por significância desejada, tendo como base os valores de semivariância obtidos através da análise de dados aleatórios. A Figura 5-5 apresenta o histograma dos dados, probabilidade acumulada e uma indicação do valor correspondente ao intervalo de confiança de 95%.

Tabela 5-13: Valores críticos, por significância desejada, para os valores de semivariância – Centro, São Paulo – SP

Intervalo de confiança	z	p-valor (α)	Limite	Limite (dados padronizados)
80%	-0,845	0,20	8,071	0,973
85%	-1,035	0,15	8,021	0,967
90%	-1,290	0,10	7,954	0,959
95%	-1,645	0,05	7,861	0,948
99%	-2,330	0,01	7,682	0,926

Figura 5-5: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Centro, São Paulo – SP

5.2.6 Padronização das semivariâncias

Tendo em vista que o valor médio da semivariância dos dados corresponde a 8,292, é possível realizar a padronização dos dados. Assim, as novas equações para a direção principal e ortogonal são apresentadas a seguir, através da Equação 5.7 e Equação 5.8.

$$\gamma(h) = 0,526 + 0,474 \left[1 - \exp \left(-\frac{h}{0,200} \right) \right] \quad (5.7)$$

$$\gamma(h) = 0,526 + 0,474 \left[1 - \exp \left(-\frac{h}{0,100} \right) \right] \quad (5.8)$$

Dessa forma, obtiveram-se os valores apresentados na Tabela 5-14, sendo que os valores de semivariância teórica padronizada para a direção principal correspondem ao indicador proposto SIVAR-G.

Tabela 5-14: Semivariâncias padronizadas pelo patamar de semivariância dos dados aleatórios, o indicador proposto corresponde aos dados teóricos da direção principal – Centro, São Paulo – SP

Vizinhança	Distância (h)	Semivariância (γ)			
		Experimental		Teórico	
		Direção Principal (0°)	Direção Ortogonal (90°)	Direção Principal (0°)	Direção Ortogonal (90°)
1	0,044	0,648	0,729	0,620	0,662
2	0,088	0,686	0,811	0,696	0,783
3	0,133	0,753	0,852	0,756	0,860
4	0,177	0,842	0,860	0,804	0,910
5	0,221	0,823	0,941	0,843	0,942
6	0,265	0,857	0,972	0,874	0,963
7	0,310	0,889	0,970	0,899	0,976
8	0,354	0,915	0,961	0,919	0,985
9	0,398	0,945	0,988	0,935	0,990
10	0,442	0,979	1,000	0,948	0,994

5.2.7 Teste de Significância

Utilizando a Tabela 5-13 de valores críticos por significância desejada, é possível analisar os dados obtidos na Tabela 5-14, que apresenta o indicador global proposto para as dez primeiras vizinhanças. Os resultados são apresentados na Tabela 5-15.

Tabela 5-15: Teste de significância (95% de confiança) para os indicadores propostos – Centro, São Paulo – SP

Vizinhança	Distância (h)	Direção Principal (0°)	H0
1	0,044	0,620	Rejeita
2	0,088	0,696	Rejeita
3	0,133	0,756	Rejeita
4	0,177	0,804	Rejeita
5	0,221	0,843	Rejeita
6	0,265	0,874	Rejeita
7	0,310	0,899	Rejeita
8	0,354	0,919	Rejeita
9	0,398	0,935	Rejeita
10	0,442	0,948	Aceita

Para 95% de confiança, considera-se que existe padrão espacial até a nona vizinhança.

5.2.8 Comparação com indicadores globais conhecidos

A Tabela 5-16, a seguir, apresenta a relação entre o índice proposto e o índice de Moran, de forma a validar os resultados encontrados.

Tabela 5-16: Comparação de valores do indicador, z e status da hipótese nula (para significância de 95%) para índice I de Moran e indicador proposto – Centro, São Paulo – SP

Viz,	Dist,	Moran				Indicador Proposto			
		Índice	Z	p-valor	H0	Índice	Z	p-valor	H0
1	0,044	0,347	20,273	0,000	Rejeita	0,620	-12,017	0,000	Rejeita
2	0,088	0,234	26,101	0,000	Rejeita	0,696	-9,633	0,000	Rejeita
3	0,133	0,171	28,780	0,000	Rejeita	0,756	-7,722	0,000	Rejeita
4	0,177	0,148	28,217	0,000	Rejeita	0,804	-6,190	0,000	Rejeita
5	0,221	0,123	27,130	0,000	Rejeita	0,843	-4,962	0,000	Rejeita
6	0,265	0,106	26,589	0,000	Rejeita	0,874	-3,978	0,000	Rejeita
7	0,310	0,087	24,564	0,000	Rejeita	0,899	-3,189	0,001	Rejeita
8	0,354	0,074	23,293	0,000	Rejeita	0,919	-2,556	0,005	Rejeita
9	0,398	0,064	21,921	0,000	Rejeita	0,935	-2,049	0,020	Rejeita
10	0,442	0,053	20,044	0,000	Rejeita	0,948	-1,643	0,050	Aceita

O valor do índice I de Moran para os dados de média de viagens por modo automóvel corresponde a aproximadamente 0,347 para a primeira vizinhança, refletindo uma autocorrelação moderada. Os valores do índice decrescem à medida que a distância considerada aumenta, todos eles são rejeitados para 95% e 99% de confiança no teste de hipótese de Moran, ou seja, há autocorrelação espacial entre os dados, ainda que moderada e fraca, para vizinhanças maiores.

O indicador SIVAR-G possui um comportamento similar ao índice de Moran. Utilizando os limites obtidos na Tabela 5-13 para 95% de confiança, rejeita-se a hipótese da variável ser espacialmente aleatória em nove das dez vizinhanças analisadas, similar ao índice I de Moran.

Como ilustrado na Figura 5-6, nesse exemplo é ainda mais claro que os índices são complementares, haja vista que o Moran mede similaridades enquanto o SIVAR-G mede dissimilaridades. Valores que correspondem à alta autocorrelação espacial encontram-se próximos a um para o Moran e zero para o SIVAR-G. Os valores obtidos para o índice proposto e de Moran quase se complementam de forma a totalizar o valor unitário.

A anisotropia deste caso não é tão forte como a observada na Cidade Fictícia, de forma a deixar mais evidente a complementariedade entre o SIVAR-G e o Moran. .

Figura 5-6: Comparação entre os índices de Moran e indicador proposto e exposição do limite crítico para o nível de confiança de 95% – Centro, São Paulo – SP

5.3 Vantagens e desvantagens do SIVAR-G

O indicador SIVAR-G apresenta desempenho similar ao Índice de Moran, contudo a interpretação dos resultados dos indicadores é inversa, já que o primeiro se baseia na dissimilaridade entre os dados e o segundo na similaridade entre os dados.

O indicador proposto é capaz de mensurar a associação espacial de dados binários sem restrições a respeito do tipo e distribuição dos dados, o que corresponde a uma limitação do Índice de Moran. Dessa forma, o indicador SIVAR-G constitui um método aplicável tanto a dados quantitativos quanto qualitativos ordinais.

Além disso, o indicador proposto mostra-se sensível à anisotropia, tornando possível investigar a associação espacial segundo direções. Essa propriedade torna o indicador capaz de uma análise mais detalhada do banco de dados, como observado na aplicação da Cidade Fictícia.

Entretanto, o indicador SIVAR-G é capaz de medir apenas associação espacial positiva, enquanto o Índice de Moran mede associação espacial positiva e negativa.

6. RESULTADOS E DISCUSSÕES – INDICADOR LOCAL

Este capítulo apresenta os resultados e discussões relativos ao indicador espacial local (SIVAR-L) para os dois bancos de dados utilizados neste trabalho. O SIVAR-L equivale aos valores das semivariâncias correspondentes aos variogramas teóricos por cada observação. Tal indicador é um desdobramento do global (SIVAR-G), a exemplo do Moran e LISA. Foram calculados e ajustados, localmente, variogramas associados a todos os pontos pertencentes aos bancos de dados.

6.1 Cidade fictícia

6.1.1 Padronização de distâncias e determinação da primeira vizinhança

Tal como no caso do indicador global, para obtenção do indicador local as distâncias foram padronizadas e a primeira vizinhança considerada equivalente àquela obtida no cálculo do SIVAR-G. Dessa forma, utilizou-se o valor de 0,032 como distância entre vizinhanças.

6.1.2 Definição dos parâmetros para cálculo dos semivariogramas experimentais

Como descrito no Capítulo 4, a análise local é feita omnidirecionalmente para cada ponto, de forma a considerar quaisquer tendências espaciais que os pontos apresentem. Adotou-se a tendência omnidirecional para o indicador local com o objetivo de englobar maior número de pontos (vizinhos) associados a cada observação. O lag foi definido anteriormente como 0,032 (primeira vizinhança) e a tolerância corresponde à metade do lag, 0,016.

6.1.3 Cálculo de semivariância experimentais para diversas vizinhanças

As variâncias foram calculadas através do código apresentado no Apêndice D – Exemplo de Cálculo de Indicador Local, compondo semivariogramas característicos para cada ponto.

A análise dos dados é realizada por meio de mapas que representam os valores das semivariâncias dos semivariogramas experimentais obtidos para cada uma das vizinhanças de análise. A seguir, são apresentados os dados obtidos para a primeira vizinhança.

Como apresentado anteriormente, o semivariograma mede a dissimilaridade entre os dados. Quanto mais parecidos forem os vizinhos de um ponto, menor o valor do indicador. Por outro lado, quanto mais diferentes forem seus vizinhos, maior o valor do indicador. Esse fenômeno pode ser observado claramente na Figura 6-1, a seguir.

Figura 6-1: Semivariância Experimental para a primeira vizinhança resultante de análise local – Cidade Fictícia

Comparando os valores do banco de dados original, apresentados na Figura 4-1, com os resultados dos semivariogramas experimentais, apresentados na Figura 6-1, é possível interpretar melhor os dados.

Na Figura 4-1 observa-se que o banco de dados apresenta uma concentração do modo ônibus (0) para os dados centrais e uma concentração do modo automóvel (1) para dados mais afastados do centro, correspondendo aos extremos leste e oeste. Segundo essa figura, para as regiões onde os dados são homogêneos (vizinhos de mesmo valor), tanto para modo ônibus (0) quanto para modo automóvel (1), observa-se baixa dissimilaridade, refletida por valores baixos de semivariância experimental, pontos brancos apresentados na Figura 6-1. Em muitas

regiões, todos os vizinhos considerados possuem o mesmo valor, tendo semivariância experimental correspondente a zero.

No entanto, nas regiões de transição (ver Figura 4-1), ou seja, onde os pontos analisados possuem vizinhos de modo ônibus (0) e de modo automóvel (1), a dissimilaridade entre o conjunto de dados é maior, retornando valores diferentes de zero para os semivariogramas experimentais (ver Figura 6-1) – pontos em escalas de cores apresentados no mapa.

6.1.4 Modelagem dos semivariogramas teóricos

Para a modelagem dos semivariogramas teóricos, por observação, utilizou-se a função *nlsLM*, mínimos quadrados não lineares. Foram testados os principais modelos utilizados para semivariogramas: esférico, exponencial e gaussiano, sendo escolhidos aqueles que apresentavam maior correlação entre os dados observados e estimados por ponto.

Aplicados os modelos aos dados experimentais, esses podem convergir, retornando os parâmetros adequados para a equação, ou não convergir, retornando uma mensagem de erro. A convergência é dada através da comparação da imprecisão das estimativas dos parâmetros com o resíduo da soma dos quadrados. Caso nenhum desses modelos convergisse, os dados foram representados por um modelo efeito pepita puro, ou seja, foram considerados aleatórios espacialmente.

Dessa forma, é possível apresentar os dados em termos de modelos de semivariogramas teóricos (Figura 6-2) utilizados e correlação dos mesmos entre valores observados e estimados (Figura 6-3), diagnosticando a qualidade da modelagem realizada.

Figura 6-2: Modelos Teóricos – Cidade Fictícia

Figura 6-3: Correlação entre valores observados e estimados pelos modelos teóricos – Cidade Fictícia

A Figura 6-2 apresenta os modelos obtidos localmente, o *software* R convergiu para um dos três modelos apresentados para 95% dos dados, 5% foram considerados efeito pepita puro, portanto sem autocorrelação espacial. Observou-se uma predominância do modelo gaussiano para os dados de regiões com valores homogêneos de modo ônibus (0) ou automóvel (1) (ver Figura 4-1). Nas regiões de transição de valores foram adotados os demais modelos, sem predominância clara.

Na Figura 6-3 é possível observar a correlação dos dados observados com os dados obtidos através dos modelos locais utilizados. Quanto maior a correlação, maior a semelhança entre os valores reais e estimados. O modelo efeito pepita puro corresponde à média das semivariâncias experimentais de um ponto, como o desvio padrão entre os dados observados e a média corresponde a zero sua correlação também resulta em zero.

Observam-se valores nulos de correlação para as regiões atribuídas ao modelo efeito pepita puro. A maioria dos demais pontos, os quais foram atribuídos outros modelos, apresentam correlação média ou alta, indicando um bom ajuste dos semivariogramas teóricos obtidos.

Aplicando os modelos, é possível obter as semivariância teóricas para todas as vizinhanças até um limite de corte de distância. O patamar obtido para cada uma das observações é diferente. Nesse formato eles não são comparáveis entre si. Na subseção 6.1.6 será apresentada a padronização adequada dos dados para que comparações, entre as observações, sejam viáveis. A Figura 6-4 apresenta os resultados de semivariância teórica para a primeira vizinhança.

Figura 6-4: Semivariância Teórica para a primeira vizinhança resultante de análise local – Cidade Fictícia

Observa-se uma concentração de baixos valores de semivariância teórica para regiões homogêneas e de altos valores para regiões de transição. Vale ressaltar que os valores de semivariâncias teóricas obtidos para a primeira vizinhança correspondem a valores baixos para todo o conjunto de observações, sendo o limite superior correspondente a 0,303.

Nas regiões homogêneas são predominantes valores baixos de semivariância, indicando que os valores dessas regiões são muito próximos entre si.

Nas regiões de transição observa-se a maioria dos valores de semivariâncias teóricas variando de 0,061 a 0,301. Esse resultado corresponde ao esperado, como essa região possui vizinhos tanto com valores zero quanto um, espera-se que existam maiores dissimilaridades entre os dados.

6.1.5 Elaboração do Teste de Hipótese

6.1.5.1 Simulação

O teste de hipótese utilizado é similar ao aplicado ao indicador global, sendo realizadas simulações de dados aleatórios com mesmas coordenadas do banco de dados original. Os

valores utilizados correspondem a números aleatórios zero ou um. O código utilizado na simulação é apresentado no Apêndice D – Exemplo de Cálculo de Indicador Local.

Foram realizadas 100 simulações, nas quais os dados foram analisados omnidirecionalmente e representados por modelos efeito pepita puro, devido à aleatoriedade espacial dos dados.

6.1.5.2 Distribuição Populacional

Foram utilizadas as semivariâncias teóricas, referentes ao efeito pepita puro. Esses valores foram analisados em conjunto, com um número médio representativo dos patamares das simulações.

Tabela 6-1: Medidas descritivas das semivariâncias teóricas para dados aleatórios de escolha modal – Cidade Fictícia

Medidas descritivas	
Média	0,246
Desvio Padrão	2,827e-2
Valor Máximo	0,255
Valor Mínimo	0,234

Aplicou-se o teste de normalidade *Kolmogorov-Smirnov* ao novo conjunto de dados, os valores obtidos correspondem a D de 0,027 e p-valor de 0,781. Assim, conclui-se que, para um nível de confiança de 95%, a distribuição analisada é similar à distribuição normal.

6.1.5.3 Pontos críticos

Partindo do princípio de que a distribuição é similar à normal, podem-se obter os valores críticos para os níveis de significância desejados, conforme descrito anteriormente no Capítulo 4. A Figura 6-5 apresenta o histograma das semivariâncias teóricas, sua curva normal correspondente e o valor crítico para confiança de 95%. A Tabela 6-2 apresenta os valores críticos calculados para os demais níveis de confiança.

Tabela 6-2: Valores críticos por significância desejada para os valores de semivariância – Cidade Fictícia

Intervalo de confiança	z	p-valor (α)	Limite	Limite (dados padronizados)
80%	-0,845	0,200	0,223	0,903
85%	-1,035	0,150	0,217	0,881
90 %	-1,290	0,100	0,210	0,852
95%	-1,645	0,050	0,200	0,811
99%	-2,330	0,010	0,181	0,733

Figura 6-5: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Cidade Fictícia

6.1.6 Padronização das variâncias teóricas

Na Tabela 6-1, observa-se que o valor médio dos patamares obtidos para os dados aleatórios corresponde a 0,246. Esse valor corresponde ao patamar máximo que os semivariogramas teóricos dos pontos podem alcançar.

Os pontos, atribuídos com o modelo efeito pepita puro, correspondem exatamente ao patamar de aleatoriedade, portanto, para todas as vizinhanças, seu indicador corresponde ao valor um, o máximo possível.

Na Figura 6-6 observam-se os resultados correspondentes às semivariâncias teóricas padronizadas, o indicador SIVAR-L, para a primeira vizinhança. A análise conjunta de vizinhanças possibilita compreender a evolução do modelo de associação espacial de cada ponto com o aumento da distância de análise. As demais vizinhanças têm os seus dados apresentados no Apêndice E – Mapas correspondentes aos semivariogramas teóricos .

Figura 6-6: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a primeira vizinhança resultante de análise local – Cidade Fictícia

Comparando a Figura 6-6 à Figura 6-2, observa-se que as regiões nas quais o modelo adotado foi efeito pepita puro correspondem a pontos onde a semivariância teórica padronizada corresponde a altas dissimilaridades.

As demais observações, as quais foram ajustados outros modelos de semivariogramas teóricos, possuem valores do indicador proposto predominantemente baixos para a primeira vizinhança, indicando alta autocorrelação espacial.

6.1.7 Teste de Significância

Comparando os valores do indicador proposto, apresentados na subseção 6.1.6 e os pontos críticos obtidos na subseção 6.1.5.3, é possível realizar as considerações de significância do indicador ponto a ponto.

Na Figura 6-7, a seguir, são apresentados os pontos significativos ou não significativos para o nível de 95% de significância, para os dados obtidos na primeira vizinhança. As demais vizinhanças têm os seus dados apresentados no Apêndice E – Mapas correspondentes aos semivariogramas teóricos.

Figura 6-7: Significância dos valores de SIVAR-L para a primeira vizinhança resultante – Cidade Fictícia

Novamente, observa-se que as regiões nas quais o modelo adotado foi efeito pepita puro os resultados são não significativos. Nos demais pontos a grande maioria acusou significância para 95% de confiança.

6.1.8 Comparação com indicadores locais comumente utilizados na literatura

O indicador proposto foi comparado com o índice Moran Local, ou LISA, de forma a compreender as semelhanças e diferenças obtidas na análise de dados através dos dois procedimentos.

O indicador LISA pode ser apresentado na forma de valores de correlação, os quais não apresentam limites de valores, e significância correspondente. O indicador é apresentado mais comumente relacionando o valor da variável do ponto em análise com o valor médio da mesma variável para seus vizinhos. A seguir são apresentados os mapas para cada uma dessas representações a respeito da primeira vizinhança.

Figura 6-8: Correlação do Indicador LISA para a primeira vizinhança – Cidade Fictícia

Figura 6-9: Significância do Indicador LISA para a primeira vizinhança – Cidade Fictícia

Figura 6-10: Mapa de *Clusters* LISA – Cidade Fictícia

As demais vizinhanças têm os seus dados apresentados no Apêndice F – Mapas correspondentes ao índice LISA.

Segundo indicador LISA, as regiões de valores homogêneos (observações e vizinhos com mesmo valor, 0 ou 1) apresentam altos valores de correlação e são significativas, enquanto as regiões de transição (observações e vizinhos com valores diferentes) apresentam baixos valores de correlação e não são significativas, como esperado.

A Figura 6-10 apresenta a análise de *cluster* dos dados. Os dados estão divididos nas categorias: valor do ponto acima da média e valores dos vizinhos também acima da média (Alto - Alto), valor do ponto abaixo da média e valores dos vizinhos também abaixo da média (Baixo - Baixo), valor do ponto abaixo da média e valores dos vizinhos acima da média (Baixo - Alto) e valor do ponto acima da média e valores dos vizinhos abaixo da média (Alto - Baixo). Através dos *clusters*, define-se claramente as regiões homogêneas de valor zero (Baixo - Baixo) e valor um (Alto-Alto) e as regiões de transição, não significativas, sem estrutura espacial observável ou identificadas por (Alto-Baixo) e (Baixo-Alto).

Comparando-se os resultados dos indicadores, observa-se claramente a diferença de interpretação dos dados. Enquanto o indicador proposto baseia-se na dissimilaridade dos dados, o indicador LISA, baseia-se na similaridade. Além disso, o intervalo de análise do LISA

compreende valores negativos (associação espacial negativa) e positivos (associação espacial positiva), enquanto o indicador proposto identifica apenas correlação espacial positiva, não sendo capaz de captar os pontos de LISA negativo, em azul da Figura 6-8.

Tanto o índice LISA quanto o SIVAR-L apresentam delimitações claras entre regiões homogêneas e de transição. A interpretação do SIVAR-L é similar aos resultados de correlação do índice LISA, atribuindo valores baixos e altos, respectivamente, para regiões de autocorrelação espacial positiva. As regiões identificadas com autocorrelação espacial negativa pelo índice LISA, o indicador SIVAR-L retorna ausência de associação espacial, para ambos os índices, essas regiões são não significativas.

Observando os resultados da primeira vizinhança (Figura 6-6) e a evolução do valor dos indicadores para as demais vizinhanças (Apêndice E – Mapas correspondentes aos semivariogramas teóricos), observa-se que o indicador SIVAR-L parte de valores baixos para as regiões homogêneas e medianos para as regiões de transição, chegando até o patamar de aleatoriedade ou próximo disso até a décima vizinhança considerada.

Da mesma forma que realizado no indicador global, é possível fazer uma comparação entre o valor do indicador proposto e do índice LISA para cada uma das vizinhanças analisadas, de forma a visualizar o comportamento dos indicadores para um mesmo ponto. Para isso foram escolhidos alguns pontos para comparação, apresentados na Figura 6-11.

Figura 6-11: Pontos selecionados para comparação – Cidade Fictícia

Os gráficos representativos da associação espacial por vizinhança desses pontos são apresentados a seguir.

Na Figura 6-12, tanto o índice LISA quanto o indicador SIVAR-L acusam alta correlação espacial que diminui com o aumento da distância considerada.

Figura 6-12: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 204 – Cidade Fictícia

Observando-se a Figura 6-13, verifica-se que para o ID 283, os índices possuem comportamentos similares ao encontrado no ID 204, em ambos os casos, o indicador SIVAR-L possui um comportamento gaussiano, sugerindo uma alteração gradual e suave na semivariância ao redor desse ponto.

Figura 6-13: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 283 – Cidade Fictícia

Na Figura 6-14 o índice LISA e o indicador proposto apresentam comportamento condizente, enquanto um decresce o outro cresce.

Figura 6-14: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 364 – Cidade Fictícia

Na Figura 6-15 sugere-se igual comportamento entre os indicadores, o indicador SIVAR-L aponta alta dissimilaridade entre os dados, enquanto o índice LISA indica baixa similaridade, alcançando valores negativos.

Figura 6-15: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 390 – Cidade Fictícia

Na Figura 6-16, o comportamento dos índices é condizente, porém o indicador proposto acusa que não há qualquer associação espacial a partir da terceira vizinhança, enquanto o LISA retorna que há uma associação baixa a partir da mesma vizinhança.

Figura 6-16: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 416 – Cidade Fictícia

Nas observações analisadas o comportamento do indicador SIVAR-L mostra-se condizente com o encontrado para o índice LISA.

6.2 Centro, São Paulo – SP

6.2.1 Padronização de distâncias e determinação da primeira vizinhança

Conforme descrito anteriormente, as primeiras etapas metodológicas para cálculo do SIVAR-L para os dados da região central de São Paulo foram a padronização das coordenadas geográficas e determinação da primeira vizinhança, tal como realizado para obtenção do indicador global. Para a primeira vizinhança foi utilizado um valor de 0,044.

6.2.2 Definição dos parâmetros para cálculo dos semivariogramas experimentais

Replicando o método utilizado na obtenção do SIVAR-L para a cidade fictícia, foram calculadas, para a região central de São Paulo, as semivariâncias experimentais omnidirecionais ponto a ponto, definindo-se apenas os valores de lag, correspondente a 0,044, e a tolerância de lag, correspondente a metade do lag, 0,022.

6.2.3 Cálculo de semivariância experimentais para diversas vizinhanças

Para o indicador local, as semivariâncias experimentais locais são calculadas ponto a ponto e não existem bibliotecas prontas no *software R*. Dessa forma, o procedimento de cálculo é descrito no Apêndice D – Exemplo de Cálculo de Indicador Local, onde constam os códigos utilizados para a aplicação do indicador.

A seguir, na Figura 6-17, são apresentados os dados relativos aos semivariogramas experimentais, obtidos para a primeira vizinhança.

Figura 6-17: Semivariância Experimental para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP

A Figura 6-17 apresenta, em sua maioria, regiões com baixa semivariância experimental, onde há baixa dissimilaridade entre os dados, ou seja, dados mais parecidos entre si. Algumas regiões são identificadas com alta dissimilaridade entre os dados, onde observações e seus vizinhos possuem valores diferentes. São exemplos de alta dissimilaridade as zonas: Jardim Bonfiglioli, Jardim Caxingui, Vila Sonia; Francisco Matarazzo, Água Branca; Independência; Parque Ibirapuera; Jurubatuba; Cupecê; Jabaquara e Cidade Vargas, como apresentado na Figura 6-18.

Figura 6-18: Semivariância Experimental para a primeira vizinhança resultante de análise local, com as regiões de baixa associação espacial em destaque – Centro, São Paulo – SP

6.2.4 Modelagem dos semivariogramas teóricos

A modelagem dos semivariogramas teóricos foi realizada de forma análoga à descrita na subseção 6.1.4. Na Figura 6-19 e na Figura 6-20, a seguir, são apresentados os tipos de modelos obtidos e a correlação entre valores observados e estimados a partir dos mesmos modelos, de forma a mensurar sua qualidade.

Figura 6-19: Modelos Teóricos – Centro, São Paulo – SP

Figura 6-20: Correlação entre valores observados e estimados pelos modelos teóricos – Centro, São Paulo – SP

Cerca de 63% dos dados obtiveram um modelo que representa sua associação espacial em detalhes, os demais 37% correspondem ao modelo efeito pepita puro, tendo autocorrelação espacial nula.

As observações, as quais foram atribuídas o modelo efeito pepita puro, correspondem a regiões próximas dos valores de alta semivariância experimental, apresentados na Figura 6-17, e outras áreas que não apresentam relação com as mesmas.

A Figura 6-20 apresenta a correlação entre valores observados e estimados pelos modelos teóricos adotados. Como observado anteriormente, regiões onde adotou-se o modelo efeito pepita puro possuem correlação nula. Os demais pontos, aos quais foram atribuídos outros modelos, apresentam em sua maioria alta correlação espacial, indicando um bom ajuste dos modelos adotados.

Aplicando os modelos, é possível obter a semivariância teórica (Figura 6-21), a qual deverá ser padronizada para corresponder ao indicador SIVAR-L

Figura 6-21: Semivariância Teórica para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP

Comparando a semivariância experimental, Figura 6-17, com a semivariância teórica, Figura 6-21, observa-se que os resultados são condizentes entre si. A maioria das regiões que apresentaram baixa semivariância teórica correspondem a áreas com vizinhos similares e

transições graduais de valores (ver Figura 4-3). Enquanto regiões com alta semivariância teórica (regiões periféricas) possuem vizinhos com mudanças abruptas de valores (ver Figura 4-3).

Sugere-se que, regiões que apresentam alta semivariância experimental (baixa similaridade entre os dados), convergem para modelos efeito pepita puro, indicando ausência de associação espacial. Esses modelos resultam em altos valores de semivariância teórica, como pode ser observado comparando a Figura 6-17 com a Figura 6-21.

6.2.5 Elaboração do Teste de Hipótese

6.2.5.1 Simulação

O teste de hipótese utilizado compreende um teste de pseudo-significância e foi realizado como descrito no Capítulo 4. Os valores utilizados correspondem a números aleatórios entre zero e dez, limites da variável média de viagens por automóvel por domicílio por quadrícula de 500 por 500 m. O código utilizado na simulação é apresentado no Apêndice D – Exemplo de Cálculo de Indicador Local.

Foram realizadas 100 simulações, nas quais os dados foram analisados omnidirecionalmente e representados por modelos efeito pepita puro, devido à aleatoriedade dos dados.

6.2.5.2 Distribuição Populacional

Assim como nos demais testes de hipótese realizados, foram utilizadas apenas as semivariâncias teóricas representadas pelo patamar dos modelos efeito pepita puro. Esses valores foram analisados em conjunto, com um número médio representativo dos patamares das simulações. Os resultados de média, desvio padrão, valores máximos e mínimos são apresentados na Tabela 6-3.

Tabela 6-3: Medidas descritivas das semivariâncias teóricas para dados aleatórios de média de viagens por automóvel – Centro, SP – SP

Medidas descritivas	
Média	8,309
Desvio Padrão	3,744
Valor Máximo	9,476
Valor Mínimo	7,174

Aplicou-se o teste de normalidade *Kolmogorov-Smirnov* ao novo conjunto de dados, os valores obtidos correspondem a D de 0,023 e p-valor de 0,713. Assim, conclui-se que para um nível de confiança de 95% a distribuição analisada é similar à distribuição normal.

6.2.5.3 Pontos críticos

Como a distribuição é similar à normal, podem-se obter os valores críticos para os níveis de significância desejados. A Figura 6-22 corresponde à representação do histograma de semivariâncias teóricas analisadas, respectiva distribuição similar à normal e delimitação do nível de confiança 95%. A Tabela 6-4 indica os valores críticos para os demais níveis de confiança.

Tabela 6-4: Valores críticos por significância desejada para os valores de semivariância – Centro, SP – SP

Intervalo de confiança	z	p-valor (α)	Limite	Limite (dados padronizados)
80%	-0,845	0,200	5,145	0,619
85%	-1,035	0,150	4,434	0,534
90%	-1,290	0,100	3,479	0,419
95%	-1,645	0,050	2,150	0,259
99%	-2,330	0,010	-0,400	-0,048

Como se pode observar, o valor correspondente a 99% de confiança é negativo, não sendo possível de ser alcançado pela distribuição em questão.

Figura 6-22: Histograma dos dados de semivariância teórica comparado à distribuição normal e linha tracejada representando o valor crítico para 95% de confiança – Centro, SP – SP

6.2.6 Padronização das semivariâncias teóricas

Na Tabela 6-3 obtém-se que o valor médio dos patamares obtidos para os dados aleatórios corresponde a 8,309, valor que corresponde ao patamar máximo que os semivariogramas teóricos dos pontos podem alcançar.

Realizando-se a padronização das semivariâncias teóricas, é possível obter a Figura 6-23, a qual apresenta os valores dos indicadores SIVAR-L para a primeira vizinhança. As demais vizinhanças têm os seus dados apresentados no Apêndice E – Mapas correspondentes aos semivariogramas teóricos .

Figura 6-23: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP

Na Figura 6-23 observa-se que as regiões nas quais foi adotado o modelo efeito pepita puro correspondem ao patamar de aleatoriedade máximo, o valor um, refletindo ausência de associação espacial.

Nas áreas de baixo indicador SIVAR-L, as quais correspondem a vizinhos com baixa dissimilaridade, são observados “bolsões” de associação espacial. Essas regiões referem-se às zonas: CEASA; Vila Anastácio; Luz, Parque Dom Pedro; Clínicas, Trianon, Oscar Freire e Jardins; Moema e Vila Olímpia; Paraisópolis; Jardim Vitória Régia e Vila Suzana; Granja Julieta; Congonhas; entre outras, como pode ser observado na Figura 6-24.

Figura 6-24: Indicador SIVAR-L, para a primeira vizinhança resultante de análise local, com os “bolsões” de associação espacial em destaque – Centro, São Paulo – SP

6.2.7 Teste de Significância

Comparando os valores do indicador proposto, apresentados na subseção anterior, e os pontos críticos obtidos na subseção 6.2.5.3, é possível determinar se os valores obtidos são significativos ou não.

Na Figura 6-25, a seguir, são apresentados os pontos significativos ou não para 95% de nível de confiança para os dados obtidos na primeira vizinhança analisada. As demais vizinhanças têm os seus dados apresentados no Apêndice E – Mapas correspondentes aos semivariogramas teóricos .

Figura 6-25: Significância das Semivariâncias Teóricas Padronizadas para a primeira vizinhança resultante de análise local – Centro, São Paulo – SP

Como comentado no item anterior, as regiões de modelo efeito pepita puro correspondem a valores aproximados pelo patamar máximo e são não significativos, enquanto os “bolsões” de valores uniformes possuem não só baixa dissimilaridade, como também são significativos para 95% de confiança.

6.2.8 Comparação com indicadores locais comumente utilizados na literatura

A seguir são apresentados os valores de correlação do índice LISA (Figura 6-26), sua significância correspondente (Figura 6-27), além do mapa de *clusters* que relaciona valores dos pontos analisados e seus vizinhos (Figura 6-28).

Figura 6-26: Correlação do Indicador LISA para a primeira vizinhança – Centro, São Paulo – SP

Figura 6-27: Significância do Indicador LISA para a primeira vizinhança – Centro, São Paulo – SP

Figura 6-28: Mapa de *Clusters* LISA – Centro, São Paulo – SP

As demais vizinhanças têm os seus dados apresentados no Apêndice F – Mapas correspondentes ao índice LISA.

O índice LISA possui muitas regiões de alto coeficiente de correlação e significativas, sugerindo a presença de “bolsões” de associação espacial. Esses “bolsões” são demarcados entre regiões de baixa correlação espacial, não significativas.

Comparando o indicador proposto SIVAR-L com o índice LISA, observa-se que ambos acusam a existência das regiões de associação espacial centrais, a nordeste, a noroeste, leste e a oeste. A região sul apresenta considerações divergentes dos indicadores. Em alguns dos pontos identificados como bolsões pelo LISA, mas não identificados pelo SIVAR-L o modelo utilizado corresponde ao efeito pepita puro, dessa forma, todos os dados dessas regiões foram aproximados para o patamar de aleatoriedade. A atribuição do modelo efeito pepita puro está relacionado ao critério de convergência da função *nlsLM* (biblioteca *minpack.lm*) do *software* R.

O Quadro 6-5 apresenta as similaridades e divergências entre a identificação de bolsões através do indicador SIVAR-L e do índice LISA.

Quadro 6-5: Zonas (nome e numeração) identificadas pelo índice LISA e pelo indicador SIVAR-L para a primeira vizinhança, a direção relativa corresponde à direção do recorte

Região	LISA	SIVAR-L	Direção relativa ao recorte
1	CEASA (105), Vila Anastácio (97)	CEASA (105), Vila Anastácio (97)	Noroeste
2	Francisco Matarazzo (92), Água Branca (93), PUC (83)	Lapa (99)	Norte
3	Sé (1), Parque Dom Pedro (2), República (5), Santa Efigênia (6), Luz (7)	Sé (1), Parque Dom Pedro (2), República (5), Santa Efigênia (6), Luz (7)	Nordeste
4	MASP (29), Consolação (32), Pacaembu (33), Paraíso (57), Campinas (73), Pamplona (74), Clínicas (76), Oscar Freire (77), Trianon (78)	MASP (29), Consolação (32), Pacaembu (33), Paraíso (57), Moema (64), Chácara Itaim (67), Vila Olímpia (68), Campinas (73), Pamplona (74), Jardins (75), Clínicas (76), Oscar Freire (77), Trianon (78), Jardim Paulistano (79), Jardim Europa (80)	Centro
5	Saúde (59), Planalto Paulista (60), Mirandópolis (61), Jardim Aeroporto (255), Vila Santa Catarina (256), Cidade Vargas (258)	Planalto Paulista (60)	Sudeste
6	Cupecê (260), Jurubatuba (263), Vila São Pedro (264)	-	Sul
7	Paraisópolis (299), Jardim Vitória Régia (300), Vila Suzana (301)	Granja Julieta (281), Fazenda Morumbi (297), Real Parque (298), Paraisópolis (299), Jardim Vitória Régia (300), Vila Suzana (301)	Sudoeste
8	Jardim Bonfiglioli (320)	-	Oeste

O indicador SIVAR-L e o índice LISA apresentam diversas regiões de concordância e discordância, de maneira geral não existem grandes divergências.

A seguir, na Figura 6-29, são apresentados alguns pontos que comparam os valores do indicador proposto e do índice LISA para as vizinhanças analisadas.

Figura 6-29: Pontos selecionados para comparação – Centro, São Paulo – SP

Os gráficos representativos da associação espacial por vizinhança desses pontos são apresentados a seguir.

Na Figura 6-30, enquanto o índice proposto não acusa associação espacial, o índice LISA identifica associação espacial positiva e mediana para a primeira vizinhança, a qual decai e torna-se negativa e fraca, equivalente aos valores do indicador proposto. Nesse exemplo pode-se observar como o indicador proposto não identifica associação espacial negativa, interpretando ausência de associação espacial e associação espacial negativa da mesma forma.

Figura 6-30: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 196 – Centro, São Paulo – SP

Na Figura 6-31, o índice LISA sugere alta autocorrelação espacial para as primeiras vizinhanças, e a estabilização em valores baixos, enquanto que o indicador proposto indica média autocorrelação espacial para as primeiras vizinhanças e então evolui para nenhuma associação espacial. Nenhum dos valores do indicador SIVAR-L é significativo.

Figura 6-31: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 229 – Centro, São Paulo – SP

Na Figura 6-32, para o ID 301 a relação entre os indicadores analisados mostra-se condizente com o tipo de estimativa realizada por cada indicador. Vale ressaltar que, para o

indicador proposto, apenas o resultado das quatro primeiras vizinhanças é significativo, enquanto que o índice LISA é significativo até a sexta vizinhança.

Figura 6-32: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 301 – Centro, São Paulo – SP

Na Figura 6-33 observa-se um comportamento condizente entre o índice LISA e o indicador SIVAR-L.

Figura 6-33: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 430 – Centro, São Paulo – SP

Para o ID 724 (Figura 6-34) os resultados de ambos os indicadores são condizentes, tanto o índice LISA quanto o indicador proposto não identificaram associação espacial. Para a primeira vizinhança há associação espacial negativa e fraca identificada somente pelo índice LISA.

Figura 6-34: Comparação entre os índices de Moran e indicador proposto SIVAR -L, apresentando o limite crítico do indicador proposto para o nível de confiança de 95% – Ponto ID 724 – Centro, São Paulo – SP

De maneira geral, os pontos analisados apresentaram uma interpretação similar entre os índices.

6.3 Vantagens e Desvantagens do SIVAR-L

O indicador SIVAR-L apresenta desempenho relativamente similar ao LISA, sobretudo considerando que ambos indicadores partem do pressuposto de isotropia dos dados, diferentemente da avaliação global. Entretanto, assim como o indicador global SIVAR-G, a sua interpretação baseia-se na dissimilaridade dos dados, enquanto o LISA baseia-se na similaridade.

O indicador proposto é capaz de mensurar a associação espacial local de um banco de dados e identificar “bolsões” de associação espacial.

O indicador SIVAR-L é dependente da modelagem de semivariogramas teóricos locais. A qualidade dos modelos adotados é apresentada na forma de correlação entre os dados estimados e observados.

Nos exemplos apresentados nesse trabalho, ao menos dois terços dos dados foram modelados através de funções que indicam a existência de associação espacial. As demais observações foram modeladas através do conceito de efeito pepita puro, ou seja, não possuem associação espacial e são representadas pelo patamar de aleatoriedade do semivariograma. Em poucos casos o indicador SIVAR-L denotou a presença ou ausência de associação espacial em pontos nos quais o índice LISA retornou uma informação contrária. Dessa forma, a modelagem ainda é um ponto a ser melhor trabalhado para que não restem dúvidas de interpretação. Pode-se apontar como uma restrição metodológica o critério de convergência para ajuste dos variogramas teóricos, associado ao *software* utilizado.

A necessidade do ajuste de um modelo de associação espacial local possibilita uma análise mais completa e detalhada do banco de dados, sendo possível investigar a estrutura espacial por ponto e compreender as particularidades das funções adotadas localmente. Por exemplo, a adoção de modelos gaussianos sugere que os vizinhos possuem uma alteração de valores contínua e gradual, enquanto que a adoção de modelos esféricos sugere mudanças ligeiramente abruptas de valores.

O indicador é aplicável a dados quantitativos e qualitativos ordinais de forma satisfatória.

7. CONCLUSÕES E RECOMENDAÇÕES

7.1 Síntese do trabalho e conclusões

A análise de dependência espacial de variáveis relativas a demanda por transportes precede a modelagem dos fluxos de viagens de uma região. Dessa forma, identificar satisfatoriamente as áreas de forte associação espacial interfere na qualidade das estimações de comportamento dos usuários de um sistema de transportes.

Esse trabalho de mestrado teve o propósito de desenvolver um procedimento metodológico para o cálculo de indicadores de associação espacial, baseados em conceitos variográficos. Buscou-se desenvolver uma técnica exploratória de análise espacial que aliasse a praticidade da análise através de indicadores às vantagens da análise geoestatística.

Para isso foram desenvolvidos dois indicadores, o primeiro no âmbito de análise global do banco de dados e o segundo no âmbito da análise local, realizando uma avaliação ponto a ponto. Os indicadores foram chamados de *Spatial Indicators based on Variographic Concepts*, ou SIVAR. Dessa forma, o indicador global corresponde a SIVAR-G e o indicador local a SIVAR-L.

A Geoestatística permite a avaliação da associação espacial em uma determinada direção, além da manipulação de dados qualitativos e quantitativos. Porém, um inconveniente de se utilizar semivariogramas é que muitos parâmetros dependem de especificações realizadas pelo usuário, como as tolerâncias do semivariograma e a modelagem do semivariograma teórico.

Quanto menor o número de adoção de parâmetros pelos usuários, o método pode ser replicado de uma forma mais segura. A utilização do parâmetro de suavização proveniente da técnica de Regressão Geograficamente Ponderada foi uma forma utilizada para obter o lag ideal e diminuir as escolhas que precisam ser realizadas pelo usuário.

Uma desvantagem de se utilizar a Regressão Geograficamente Ponderada para obtenção do lag (*bandwidth*) é que esse método utiliza um modelo de regressão linear como base para cálculos, assim é importante que o banco de dados analisado forneça outras informações além da variável regionalizada de interesse. Caso não existam outras variáveis o modelo poderá se basear apenas em latitude e longitude, podendo comprometer a qualidade das previsões.

É válido lembrar que a utilização do *bandwidth* da RGP é uma forma de se calcular a distância ideal de análise do conjunto de vizinhos, mas não é uma regra. Assim como para o caso do índice de Moran e LISA, o indicador proposto pode ser calculado para qualquer distância que o usuário queira, o parâmetro de suavização é apenas uma sugestão de distância com base teórica.

Nesse trabalho, os indicadores SIVAR foram aplicados a dados binários e contínuos, os resultados foram comparados aos índices de Moran global e local, ou LISA. As diferenças metodológicas entre os procedimentos são apresentadas no Quadro 7-1. A Figura 7-1 apresenta as considerações de cálculo dos indicadores tradicionais, compreendendo o índice de Moran, e as considerações de cálculo utilizadas para o indicador SIVAR.

Quadro 7-1: Comparações metodológicas entre índice de Moran e SIVAR

Características	Indicador Moran (Global e Local)	SIVAR (Global e Local)
Tipo de dados	Quantitativos, preferencialmente contínuos	Quantitativos ou Qualitativos
Consideração dos dados	(1) (Figura 7-1)	(2) (Figura 7-1)
Direção	Ominidirecional	Omnidirecional ou em uma direção específica, opcional
Base de Cálculo	Covariância	Variância
Identifica correlação espacial	Positiva e Negativa	Positiva
Modelo teórico de variação espacial	É possível, mas não é o objetivo da análise	É possível e é a base da análise variográfica
Teste de Hipótese	Pseudo-significância e baseado em permutação de valores de correlação	Pseudo-significância baseado na distribuição populacional de um banco de dados aleatório

Figura 7-1: Configuração (1) de indicadores tradicionais e (2) do indicador proposto para uma análise omnidirecional, d corresponde à distância da primeira vizinhança e a tolerância de (2) é dada por $d/2$.

Comparando indicadores tradicionais, como Moran, c de Geary e família de estatísticas $G(d)$ de Getis e Ord ao SIVAR, observa-se que os primeiros são adequados para dados quantitativos, enquanto o segundo é aplicável a dados quantitativos e qualitativos. A geoestatística apresenta a possibilidade de se trabalhar com uma maior gama de tipos de dados, inclusive dados categóricos, a partir de estruturas bifásicas ou multifásicas.

Nesse trabalho, o indicador SIVAR foi aplicado a dados contínuos (quantitativos) e binários (qualitativos). Para a análise global e local a interferência do tipo de dado foi mínima, apresentando resultados satisfatórios tanto para dados quantitativos contínuos quanto para qualitativos binários.

Em um banco de dados altamente correlacionado espacialmente, como é o caso dos dados binários da cidade fictícia, os semivariogramas teóricos locais não apresentaram

convergência para os modelos usuais em 5% dos dados, indicando que essas observações não apresentavam estrutura espacial.

Destaca-se que as aplicações dos indicadores Moran e LISA para dados binários realizadas nesse estudo contrariam as premissas desses índices. A utilização dessas técnicas constituiu uma forma de comparação e validação dos resultados do indicador SIVAR.

Os indicadores tradicionais realizam uma consideração correspondente a (1) da Figura 7-1, ou seja, do ponto de análise até a distância pré-determinada. Já o indicador proposto utiliza uma análise na configuração (2) da Figura 7-1, ou seja, como uma tolerância equivalente à metade da distância da primeira vizinhança, formando uma “coroa”.

Considerar todos os vizinhos desde a origem até uma certa distância d implica em uma maior diferença (menor similaridade, maior dissimilaridade) entre os dados analisados devido ao aumento do volume de dados e não necessariamente ao aumento da diferença entre eles. A consideração de uma “coroa” inibe, em certo grau, esse fenômeno, já que apesar da área considerada ser maior a cada vizinhança, os pontos considerados para o cálculo do indicador para uma vizinhança não são utilizados no cálculo do indicador da vizinhança seguinte.

Ao contrário dos indicadores tradicionais, o indicador proposto permite considerar uma determinada direção para a análise de dados, sendo apropriado para dados anisotrópicos. Caso exista uma tendência espacial em uma certa direção o indicador pode retornar resultados melhores de associação espacial quando comparados aos que seriam obtidos nas avaliações tradicionais omnidirecionais, enriquecendo a análise

Este fenômeno pode ser observado no banco de dados da Cidade Fictícia. Tal banco apresenta alta correlação espacial segundo o indicador proposto e o índice de Moran. Contudo, o indicador proposto foi calculado segundo a direção de maior continuidade, retornando uma associação espacial superior àquela identificada por Moran.

Um acontecimento similar não foi observado no banco de dados de São Paulo, o qual não apresentava forte tendência espacial em determinada direção do banco de dados e obteve resultados dos indicadores proposto e Moran praticamente complementares entre si, tendo a soma correspondente ao valor um.

O índice c de Geary (1954) assemelha-se ao cálculo realizado no presente trabalho para as semivariâncias experimentais, porém não considera possíveis anisotropias entre diferentes direções, realizando um cálculo omnidirecional, portanto incapaz de considerar fenômenos anisotrópicos.

O índice de Moran baseia-se em um cálculo de covariância, enquanto o indicador SIVAR baseia-se em variância. Essas diferenças não comprometem a qualidade dos cálculos,

apenas determinam as diferentes interpretações dos resultados. Assim, espera-se que os resultados dos indicadores sejam condizentes entre si, porém não sejam iguais, já que as bases de cálculo são diferentes.

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 e pequena. Já o índice de Moran permite identificar associação espacial positiva e negativa, embora para grande parte das análises apenas a associação positiva seja interessante. Pode-se afirmar que, para o indicador SIVAR valores próximos de um correspondem tanto a ausência de associação espacial quanto à associação espacial negativa.

Através dos indicadores tradicionais, é possível analisar a associação espacial dos dados para cada vizinhança, porém esse não é o objetivo da análise. Como o indicador proposto baseia-se em conceitos variográficos, a modelagem de um semivariograma teórico, que represente a associação espacial dos dados, é uma consequência da análise realizada. Dessa forma, a opção de representar a associação espacial através de um modelo torna a análise mais rica e possibilita compreender melhor a estrutura dos dados a partir do modelo adotado para representá-los.

Quando não há convergência dos modelos aplicados, adota-se um modelo efeito pepita puro, indicando ausência de estrutura espacial. Em poucos casos o indicador SIVAR-L retornou presença ou ausência de associação espacial, enquanto o índice LISA produziu uma informação contrária. Esses casos e a própria modelagem precisam ser trabalhados para que não existam dúvidas de interpretação.

O teste de hipótese utilizado para o índice Moran baseia-se em um teste de pseudo-significância, comparando a distribuição de casos aleatórios com a distribuição normal padrão. Para o índice LISA o teste de hipótese consiste na comparação com os dados permutados da população de correlações geradas para aquele banco de dados.

Para o indicador proposto optou-se por desenvolver um teste de pseudo-significância para o indicador global e local. No caso do indicador local, foram simulados dados aleatórios para todos os pontos, obtendo-se a análise da população de cada ponto. Para fins práticos utilizou-se os valores dessas análises em conjunto.

Conclui-se que o indicador SIVAR-Global é aplicável satisfatoriamente a dados contínuos (quantitativos) e binários (qualitativos), constituindo uma técnica mais abrangente que suas antecessoras, como índice de Moran, c de Geary e estatística $G(d)$ de Getis e Ord. O indicador é capaz de identificar e quantificar a associação espacial de dados espacialmente distribuídos de forma adequada e condizente com o índice de Moran. O SIVAR-G mostrou-se

sensível à direção escolhida, sendo capaz de identificar anisotropia dos dados e retornar diferentes valores de associação espacial a depender da direção analisada.

O indicador SIVAR-Local é aplicável satisfatoriamente a dados contínuos e binários. É capaz de identificar “bolsões” de associação espacial e *outliers*. Através da modelagem de semivariogramas teóricos para cada ponto, o indicador SIVAR-Local permite a identificação da estrutura espacial, localmente, através dos semivariogramas ajustados. A qualidade dos modelos adotados para os semivariogramas teóricos é avaliada através da correlação entre os dados estimados e observados.

A determinação de ambos os indicadores depende das seguintes escolhas do usuário: identificação do melhor modelo de regressão linear utilizada na Regressão Geograficamente Ponderada, determinação das tolerâncias do método de cálculo, identificação da direção principal e ortogonal (indicador global).

Ressalta-se que a geoestatística é aplicável a dados pontuais ou dados de área de igual tamanho, os quais podem ter seus dados representados na forma de um centroide.

7.2 Recomendações

Para um método ser largamente utilizado esse deve ter sua aplicação facilitada o máximo possível. Dessa forma, recomenda-se que esse método seja amparado por um código amigável, biblioteca no *software* R ou programa de livre distribuição para que possa ser aplicado pelo maior número de usuários possível.

As limitações intrínsecas à modelagem do método devem ser superadas, em especial a modelagem dos semivariogramas teóricos para o caso do SIVAR-L. Uma maior gama de modelos pode ser testada, dando maior possibilidade à convergência para modelos conhecidos e, preferivelmente, com patamares obtidos a baixas semivariâncias.

Alguns semivariogramas experimentais podem ser melhores representados através da função potência, o que indica a inexistência de patamar de aleatoriedade. Esses casos devem ser identificados e tratados, de forma a não retornar falsas respostas a respeito da presença ou não de associação espacial.

É possível realizar o teste de hipótese de maneira alternativa a elaborada. Pode-se utilizar a mesma abordagem apresentada nos índices Moran e LISA, gerando-se permutações dos valores de correlação obtidos para o conjunto de dados.

Outra possibilidade é adaptação do teste de hipótese do indicador proposto local, desenvolvendo uma análise de pseudo-significância local, sem agregar os dados para análise da

forma que foi realizada. Dessa forma, cada ponto teria seus pontos críticos específicos, resultando em maior qualidade na comparação com a distribuição normal e verificação da significância dos pontos.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L.; GRIFFITH, D. A. (1988) Do spatial effects really matter in regression analysis? **Papers in Regional Science**. v. 65, p. 11-34 DOI: 10.1111/j.1435-5597.1988.tb01155.x

ANSELIN, L. (1995) Local Indicators of Spatial Association—LISA. **Geographical Analysis**. v. 27, p. 93-115 DOI: 10.1111/j.1538-4632.1995.tb00338.x

ANSELIN, L.; BERA, A. K.; FLORAX, R.; YOON, M. J. (1996) Simple diagnostic tests for spatial dependence. **Regional Science and Urban Economics**. v. 26, p. 77-104 DOI: 10.1016/0166-0462(95)02111-6

ANSELIN, L. (2010) Thirty years of spatial econometrics. **Papers in Regional Science**. v. 89, p. 3-25 DOI: 10.1111/j.1435-5957.2010.00279.x

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.

BEN-AKIVA, M. E.; LERMAN, S. R. (1985) **Discrete choice analysis: theory and application to travel demand**. MIT press, Cambridge, p. 390. DOI: 10.1007/978-1-4615-5203-1_2

BERRY, B.J.L.; MARBLE, D.F. (1968). **Spatial analysis**. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

BIVAND, R. S.; PEBESMA, E.; GOMEZ-RUBIO, V. (2013) **Applied spatial data analysis with R**, Second edition. Springer, NY. <http://www.asdar-book.org/>

BIVAND, R. S.; YU, D. (2017). **spgwr: Geographically Weighted Regression**. R package version 0.6-31. <https://CRAN.R-project.org/package=spgwr>

BJORNSTAD, O. N. (2016). **ncf: Spatial Nonparametric Covariance Functions**. R package version 1.1-7. <https://CRAN.R-project.org/package=ncf>

BLACK, J. A.; PÁEZ, A.; SUTHANAYA, P. A. (2002) Sustainable urban transportation: Performance indicators and some analytical approaches. **Journal of Urban Planning and Development**, v. 128, p.184-209 DOI: 10.1061/(ASCE)0733-9488(2002)128:4(184)

BURROUGH, P. A. (1986) **Principles of Geographical Information Systems for Land Resources Assessment**, Oxford Press, New York, 155-165 p.

CAMARGO, E. C. G. (1998) **Geoestatística: Fundamentos e Aplicações. Geoprocessamento para projetos ambientais**. São José dos Campos: INPE.

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

CLIFF, A.; ORD, K. (1981) **Spatial processes: Models and applications**. Londres, Pion, p. 266.

COSTA, A. S. (2013) **Proposta de um método para estimação de escolha modal através da Geoestatística**. Mestrado em Engenharia Ambiental Urbana – Universidade Federal da Bahia, Salvador, 147 p.

CRESSIE, N. (1991) **Statistics for Spatial Data**. New York, Wiley. 928 p.

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.

ELZHOV, T. V.; MULLEN, K. M.; SPIESS, A. N.; BOLKER, B. (2016) **R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds**. R package version 1.2-1. <https://CRAN.R-project.org/package=minpack.lm>

FISCHER, M. M.; GETIS, A. (2009) **Handbook of applied spatial analysis: software tools, methods and applications**. Springer Science & Business Media, p. 461-467.

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

FORTIN, M. J.; DALE, M. R. T.; HOEF J. M. (2002) Spatial analysis in ecology. **Wiley StatsRef: Statistics Reference Online**. DOI: 10.1002/9780470057339.vas039.pub2

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

GOMES, V. A.; PITOMBO, C. S.; ROCHA, S. S.; SALGUEIRO, A. R. (2016) Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting. **Open Journal of Statistics**. v.6, p. 514-527 DOI: 10.4236/ojs.2016.63044

GOOVAERTS, P. (2010) Spatial uncertainty and hypothesis testing in medical geography: A geostatistical perspective. In: 9th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. **Anais...** Leicester; United Kingdom. p. 5-8

JOURNEL, A. G.; HUIJBREGTS, C. J. (1978) **Mining Geoestistics**. London: Academic Press.

JOURNEL, A. G. (1989) **Fundamentals of Geoestistic in Five Lessons**. Short Course in Geology, v.8, Washington: American Geophysical Union.

KALOGIROU, S. (2017) **lctools: Local Correlation, Spatial Inequalities, Geographically Weighted Regression and Other Tools**. R package version 0.2-6. <https://CRAN.R-project.org/package=lctools>

KELEJIAN, H. H.; ROBINSON, D. P. (1992) Spatial autocorrelation. A new computationally simple test with an application to per capita county police expenditures. **Regional Science and Urban Economics**. v. 22, p. 317-331 DOI: 10.1016/0166-0462(92)90032-V

KELEJIAN, H. H.; PRUCHA, I. R. (2001) On the asymptotic distribution of the Moran I test statistic with applications. **Journal of Econometrics**. v. 104, p. 219-257 DOI: 10.1016/S0304-4076(01)00064-1

KREMPI, A. P. (2004) **Explorando recursos de estatística espacial para análise da acessibilidade da cidade de Bauru**. Dissertação (Mestrado em Engenharia Civil) – Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 82 p.

KRIGE, D. G. (1951) A Statistical Approaches to Some Basic Mine Valuation Problems on the Witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**. v. 52, p. 119-139.

LINDNER, A. (2015) **Análise desagregada de dados de demanda por transporte através de modelagem geoestatística e tradicional**. Dissertação de Mestrado em Programa de Pós-graduação em Engenharia de Transportes – Escola de Engenharia de São Carlos Universidade de São Paulo.

LINDNER, A., PITOMBO, C. S., ROCHA, S. S., QUINTANILHA, J. A. (2016) Estimation of transit trip production using Factorial Kriging with External Drift: an aggregated data case study. **Geo-spatial Information Science**. v. 19, p. 245-254. DOI: 10.1080/10095020.2016.1260811

LINDNER, A.; PITOMBO, C. S. (2018) A Conjoint Approach of Spatial Statistics and a Traditional Method for Travel Mode Choice Issues. **Journal of Geovisualization and Spatial Analysis**, v. 2, p. 1-13. DOI: 10.1007/s41651-017-0008-0

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.

MACHIWAL, D.; MISHRA, A.; JHA, M.K.; SHARMA, A.; SISODIA, S.S. (2012) Modeling Short-Term Spatial and Temporal Variability of Groundwater Level Using Geostatistics and GIS. **Natural Resources Research**. v. 21, p. 117-136 DOI: 10.1007/s11053-011-9167-8

MATHERON, G. (1970) **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.

METRÔ (2012) Companhia do Metropolitano de São Paulo. **Pesquisa de Mobilidade 2012**. Disponível em < <http://www.metro.sp.gov.br/> > Acesso em: 10 ago. 2016.

MORAN, P.A. (1950) Notes on continuous stochastic phenomena. **Biometrika**. v. 37, p. 17-23 DOI:10.2307/2332142

NANOS, N.; TADESSE, W.; MONTERO, G.; GIL, L.; ALIA, R. (2001) Spatial stochastic modeling of resin yield from pine stands. **Canadian Journal of Forest Research**. v. 31, p. 1140-1147 DOI: 10.1139/cjfr-31-7-1140

NASS, C.; GARFINKLE, D. (1992) Localized autocorrelation diagnostic statistic (LADS) for spatial models. Conceptualization, utilization, and computation. **Regional Science and Urban Economics**. v. 22, p. 333-346 DOI: 10.1016/0166-0462(92)90033-W

ORTÚZAR, J.D.; WILLUMSEN, L.G. (2011). **Modelling Transport**. Londres: Wiley, 4ª Edição. 586 p.

PITOMBO, C.S., SOUSA, A.J. (2009) Aplicação de Conceitos Geoestatísticos para Análise de Geração de Viagens Urbanas. In: XXIII Congresso de Pesquisa e Ensino em Transportes (ANPET), Vitória. **Anais...** Vitória.

PITOMBO, C. S.; DA COSTA, A .S. G.; SALGUEIRO, A. R. (2015) Proposal of a sequential method for spatial interpolation of mode choice. **Boletim de Ciências Geodésicas**. v. 21, p. 274-289 DOI: 10.1590/S1982-21702015000200016

PITOMBO, C. S.; SALGUEIRO, A. R.; DA COSTA, A .S. G.; ISLER, C. A. (2015) A two-step method for mode choice estimation with socioeconomic and spatial information. **Spatial Statistics**. v. 11, p. 45-64. DOI: 10.1016/j.spasta.2014.12.002

R CORE TEAM (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RIBEIRO, P. J.; DIGGLE, P. J. (2016) **geoR: Analysis of Geostatistical Data**. R package version 1.7-5.2. <https://CRAN.R-project.org/package=geoR>

ROCHA, S. S., PITOMBO, C. S., SALGUEIRO, A. R. (2016) Interpolação espacial de viagens urbanas por transporte coletivo através de amostra sistemática desagregada artificial. **Revista Brasileira de Cartografia**. v. 68, ed.4.

ROCHA, S. S. ; LINDNER, A. ; PITOMBO, C. S. (2017) Proposal of a Geostatistical procedure for transportation planning field. **Boletim de Ciências Geodésicas** (impresso), v. 23, p. 636-653. DOI: 10.1590/s1982-21702017000400042

RODRIGUES DA SILVA, A. N. (2008) **Pesquisa Origem Destino da Cidade de São Carlos**. Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos – USP. São Carlos.

RUIZ, M.; LÓPEZ, F.; PÁEZ, A. (2010) Testing for spatial association of qualitative data using symbolic dynamics. **Journal of Geographical Systems**. v. 12, p. 281-309 DOI: 10.1007/s10109-009-0100-1

SARKAR, D. (2008) **Lattice: Multivariate Data Visualization with R**. Springer, New York. ISBN 978-0-387-75968-5

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

SOARES, A. (2000) **Geoestatística para as Ciências da Terra e do Ambiente**. 2. Ed. Lisboa: IST Press, 214 p.

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

TOEPFER, S.; ELLSBURY, M. M.; ESCHEN, R.; KUHLMANN, U. (2007) Spatial clustering of *Diabrotica virgifera virgifera* and *Agriotes ustulatus* in small-scale maize fields without topographic relief drift. **Entomologia Experimentalis et Applicata**. v. 124, p.61-75 DOI: 10.1111/j.1570-7458.2007.00549.x

YAMAMOTO, J. K.; LANDIM, P. M, B. (2013) **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos. 215 p.

WACKERNAGEL, H. (2003) **Multivariate Geostatistics: An introduction with applications**. 3. Ed. Berlin: Springer, 2003.

WICKHAM, H. (2009) **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York. <http://ggplot2.org>

ZHANG, C.; MCGRATH, D. (2004) Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods. **Geoderma**. v. 119, p. 261-275 DOI: 10.1016/j.geoderma.2003.08.004

9. LISTA DE ANEXOS E APÊNDICES

Anexo A – Banco de dados Centro, São Paulo – SP

Figura 9-1: Recorte realizado na região central da cidade de São Paulo

Figura 9-2: Macrozonas da Região Central de São Paulo selecionada para o estudo

Legend

Join_Output_3

<all other values>

ZONA, NOMEZONA

1, Sé	58, Bosque da Saúde	103, Bela Aliança
2, Parque Dom Pedro	59, Saúde	104, Vila Hamburguesa
3, Praça João Mendes	60, Planalto Paulista	105, CEASA
4, Ladeira da Memória	61, Mirandópolis	106, Vila Leopoldina
5, República	62, Parque Ibirapuera	222, Ipiranga
6, Santa Efigênia	63, Jardim Luzitânia	223, Sacomã
7, Luz	64, Moema	224, Alto do Ipiranga
13, João Teodoro	65, Bandeirantes	225, Vila São José
16, Brás	66, Vila Nova Conceição	226, Vila Monumento
17, Gasômetro	67, Chácara Itaim	233, Água Funda
18, Independência	68, Vila Olímpia	234, Jardim da Saúde
19, Cambuci	69, Hédio Pelegrino	235, Vila Gumerindo
20, Glicério	70, Brooklin	236, Jardim Previdência
21, Aclimação	71, Vila Cordeiro	251, Joaquim Nabuco
22, Pires da Mota	72, Berrini	252, Vieira de Moraes
23, Centro Cultural	73, Campinas	253, Campo Belo
24, Liberdade	74, Pamplona	254, Congonhas
25, Treze de Maio	75, Jardins	255, Jardim Aeroporto
26, Bexiga	76, Clínicas	256, Vila Santa Catarina
27, Bela Vista	77, Oscar Freire	257, Jabaquara
28, São Carlos do Pinhal	78, Trianon	258, Cidade Vargas
29, Masp	79, Jardim Paulistano	260, Cupecê
30, Higienópolis	80, Jardim Europa	263, Jurubatuba
31, Vila Buarque	81, Pinheiros	264, Vila São Pedro
32, Consolação	82, Vila Madalena	281, Granja Julieta
33, Pacaembu	83, PUC	282, Chácara Flora
34, FAAP	84, Cardoso de Almeida	283, Santo Amaro
35, Santa Cecília	85, Zéquinha de Abreu	284, Vila Miranda
36, Marechal Deodoro	86, Sumaré	295, Morumbi
37, Rudge	87, Perdizes	296, Joquei Clube
49, Ana Rosa	88, Vila Anglo Brasileira	297, Fazenda Morumbi
50, Jardim da Glória	89, Pompéia	298, Real Parque
51, Chácara Klabin	92, Francisco Matarazzo	299, Paraísopolis
52, Vila Mariana	93, Água Branca	300, Jardim Vitória Régia
53, Santa Cruz	94, Vila Beatriz	301, Vila Suzana
54, Vila Clementino	95, Alto de Pinheiros	306, Portal do Morumbi
55, França Pinto	96, Boaçaça	308, Vila Sonia
56, Rodrigues Alves	97, Vila Anastácio	318, Butantã
57, Paraíso	99, Lapa	319, Jardim Caxingui
	100, Vila Ipojuca	320, Jardim Bonfiglioli
	101, Alto da Lapa
	102, Gavião Peixoto	

Figura 9-3: Legenda correspondente às zonas na Figura 9-2

A Figura 9-2 e a Figura 9-3 fazem parte da pesquisa de doutorado ainda em andamento da aluna Samille Rocha, no Departamento de Transportes, EESC – USP.

Apêndice A – Regressão Linear

A etapa de regressão linear foi executada para possibilitar a aplicação do método de Regressão Geograficamente Ponderada, de forma a obter o parâmetro de suavização.

Os seguintes resultados referem-se à cidade fictícia.

Tabela 9-1: Coeficientes do modelo de regressão linear múltipla – Cidade Fictícia

Modelo	Coeficientes não normalizados		Estatística t	Significância
	B	Erro Norm.		
Tempo de viagem modo ônibus	-0,064	0,014	-4,494	0,000
Preço modo ônibus	0,370	0,051	7,222	0,000
Renda	0,040	0,013	3,163	0,002

Os seguintes resultados referem-se ao Centro da cidade de São Paulo – SP.

Tabela 9-2: Coeficientes do modelo de regressão linear múltipla – Centro, São Paulo – SP

Modelo	Coeficientes não normalizados		Estatística t	Significância
	B	Erro Norm.		
Média de Viagens Totais	0,423	0,008	53,319	0,000

Apêndice B – Semivariogramas experimentais e teóricos – Indicador Global

Os seguintes resultados referem-se à cidade fictícia.

Figura 9-4: Variogramas experimentais da variável escolha modal – Cidade Fictícia

Os seguintes resultados referem-se ao Centro da cidade de São Paulo – SP.

Figura 9-5: Semivariogramas experimentais da variável escolha modal – Centro, São Paulo – SP

Apêndice C – Exemplo de Cálculo de Indicador Global

Neste item será exposto em detalhes como calcular o indicador global de associação espacial e códigos utilizados no procedimento. O método ainda não é automático, o desenvolvimento de uma biblioteca amigável no *software R* é uma recomendação para projetos futuros.

No exemplo de cálculo são analisados os dados provenientes da Cidade Fictícia, apresentados no item 5.1.

1. Entrada dos dados

O usuário deve escolher um diretório e abrir uma planilha de dados. A planilha deve conter no mínimo as coordenadas dos pontos (X,Y) e a variável de interesse. Quando importados, os dados da planilha estarão no formato *data.frame*.

```
> setwd("C:\\Users\\claudia.naizer\\Dropbox\\Dissertacao\\_IndicadorGlobal")
> dados <- read.table("CidFicBIN.txt", h=T)
```

Considera-se que os dados de entrada já estão com as distâncias padronizadas, procedimento o qual foi executado no *software Excel*.

2. Regressão Linear

Para a aplicação da técnica de regressão geograficamente ponderada é necessário utilizar um modelo de regressão linear. Dessa forma, procura-se obter o melhor modelo possível, ou seja, com maior valor de R^2 e variáveis significativas.

```
> model <- lm (b ~ a) #exemplo de como modelos lineares são calculados no R
> model <- lm(ESCOLHA ~ TEMPO_BUS + PRECO_BUS + RENDA - 1)
> summary(model)
```

Call:

```
lm(formula = ESCOLHA ~ TEMPO_BUS + PRECO_BUS + RENDA - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.8252	-0.4552	0.1041	0.2760	0.5175

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
TEMPO_BUS	-0.06365	0.01416	-4.494	8.46e-06 ***
PRECO_BUS	0.36975	0.05120	7.222	1.63e-12 ***
RENDIA	0.04018	0.01270	3.163	0.00164 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3965 on 577 degrees of freedom

Multiple R-squared: 0.7191, Adjusted R-squared: 0.7176

F-statistic: 492.4 on 3 and 577 DF, p-value: < 2.2e-16

3. Parâmetro de Suavização da Regressão Geograficamente Ponderada

Para o cálculo do parâmetro de suavização utiliza-se a biblioteca *spgwr* (BIVAND; YU, 2017), calcula-se *bandwidth* através do critério de validação cruzada e Akaike. Como demonstrado a seguir.

```
> attach(dados)
> mapa = SpatialPointsDataFrame(data=dados,coords=cbind(X,Y))
> names(mapa)
> bwcv = gwr.sel(model, data=mapa, adapt=F, method="cv") #método de validação cruzada
> bwcv
Bandwidth: 0.03779545 CV score: 7.765348
> bwaic = gwr.sel(model, data=mapa, adapt=F, method="aic") #método de akaike
> bwaic
Bandwidth: 0.03148739 AIC: -997.8258
```

É adotado o menor valor entre eles os resultados.

4. Índice de Moran Global

O índice de Moran Global independe dos cálculos do indicador e pode ser realizado em qualquer momento do procedimento, desde que seja utilizado para comparação na etapa final. Contudo, o índice de Moran Global calculado através da biblioteca *lctools* (KALOGIROU, 2017), já possui a condição de apenas realizar o cálculo do índice de forma a evitar que mais do que 10% dos dados não tenham vizinhos.

Dessa forma, utilizando esta biblioteca e calculando o índice de Moran Global nesse momento do procedimento é possível realizar duas atividades em uma, verificar se o bandwidth obtido abrange adequadamente o banco de dados e calcular o índice de Moran Global.

```
> Coords <- cbind(dados$X, dados$Y)
> moran <- moransI.v(Coords, bws, dados$ESCOLHA, family='fixed')
> moran
```

ID	k	Moran's I	Expected I	Z resampling	P-value resampling	Z randomization	P-value randomization	
[1,]	1	0.032	0.9017513	-0.001727116	36.31555	9.194275e-289	36.25456	8.423e-288
[2,]	2	0.064	0.8503549	-0.001727116	46.46670	0.000000e+00	46.38864	0.000e+00
[3,]	3	0.096	0.8463643	-0.001727116	53.37160	0.000000e+00	53.28196	0.000e+00
[4,]	4	0.128	0.7817407	-0.001727116	65.87095	0.000000e+00	65.76036	0.000e+00
[5,]	5	0.160	0.6993759	-0.001727116	74.86733	0.000000e+00	74.74168	0.000e+00
[6,]	6	0.192	0.6434994	-0.001727116	82.70449	0.000000e+00	82.56571	0.000e+00
[7,]	7	0.224	0.5826857	-0.001727116	85.04339	0.000000e+00	84.90069	0.000e+00
[8,]	8	0.256	0.5079023	-0.001727116	83.97111	0.000000e+00	83.83029	0.000e+00
[9,]	9	0.288	0.4344246	-0.001727116	81.82379	0.000000e+00	81.68673	0.000e+00
[10,]	10	0.320	0.3600095	-0.001727116	76.02325	0.000000e+00	75.89608	0.000e+00

5. Semivariogramas experimentais

Os semivariogramas experimentais podem ser realizados através do *software SGeMS* ou através do *software R*. O *software SGeMS* permite gerar semivariogramas para várias direções, visualiza-los simultaneamente com a mesma escala e modela-los simultaneamente. O último ponto é uma vantagem em comparação com o *software R*, o qual permite a modelagem de apenas um semivariograma experimental por vez.

Utilizando o *SGeMS* os dados devem ser importados para o programa e analisados no item Data Analysis > Variogram, onde é possível inserir a quantidade de pontos de interesse, lag, tolerância do lag, direção, mergulho (não utilizado na análise 2D), tolerância angular e tolerância de banda. Gerados os semivariogramas, observa-se qual o menor patamar (maior continuidade espacial) e maior patamar (menor continuidade espacial).

No *software R* é necessário utilizar a biblioteca *geoR* (RIBEIRO; DIGGLE, 2016). Como não há cálculo automático e simultâneo, devem ser calculados os semivariogramas para cada uma das direções previstas e então plotados em conjuntos para análise de continuidade por direção, determinando-se a direção principal e ortogonal.

```
> dados <- as.geodata(dados) #transforma a informação em geográfica
> var.exp.0 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
> var.exp.30 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
> var.exp.60 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
> var.exp.90 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
> var.exp.120 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
> var.exp.150 <- variog(df, uvec=seq(0.032,1,by=0.032), direction=0,tolerance=22 ,unit.angle = "degrees")
```

Os valores de semivariância experimentais encontram-se na Tabela 5-3.

6. Semivariogramas teóricos

Como comentado, através do *software SGeMS* é possível realizar a modelagem no mesmo ambiente onde foram gerados os semivariogramas teóricos de forma simultânea em todos os gráficos.

No *software R* é necessário definir as direções principal e ortogonal e realizar a calibração independente em cada semivariograma, mas de forma que corresponda o mesmo modelo para as duas direções.

Uma sugestão de valores inicial pode ser dada através da função *variofit*, que determina o melhor modelo através de mínimos quadrados não lineares. A seguir é apresentado um exemplo para o modelo esférico.

```
> var.teor.0 <- variofit(var.exp.0, cov.model = "sph", fix.nugget=F)
> var.teor.90 <- variofit(var.exp.90, cov.model = "sph", fix.nugget=F)
```

Ou é possível aplicar a função *eyefit*, na qual o ajuste do modelo é feito de forma manual.

```
> var.teor.0 <- eyefit(var.exp.0)
> var.teor.90 <- eyefit(var.exp.90)
```

Os semivariogramas obtidos podem ser observados no Apêndice B – Semivariogramas experimentais e teóricos. Os valores de semivariância teóricos encontram-se na Tabela 5-3.

7. Teste de Hipótese

Para o teste de hipótese foi simulada uma população de mesmas coordenadas que a original, porém com dados da variável de análise aleatórios. A função desenvolvida para essa simulação é apresentada a seguir.

```
teste1 <- function(x,y,data.type,k,max){

#k é o numero de ptos
#max é o limite máximo do intervalo, o mínimo é zero

library("geoR", lib.loc="~/R/win-library/3.4")

setwd("C:\\Users\\claudia.naizer\\Dropbox\\Dissertação\\Código")
#define diretório para salvar os outputs

n<-100
matexp <- matrix(ncol=10, nrow=n) #define a matriz base que será preenchida com os dados
experimentais para 20 pontos do variograma
matteor <- matrix(ncol=1, nrow=n) #define a matriz base que será preenchida com os dados
teóricos para 1 ponto, o efeito pepita puro

for(i in 1:n){#loop para os 100 variogramas

  if (data.type=='bin'){
    z <- rep(0,k)
```

```

for (m in 1:k){
  z[m] <- sample(0:1, 1)
}
}
if (data.type=='cont'){
  z <- runif(k,0,max)
}
#define z como k números aleatórios uniformemente distribuídos entre 0 e 1 (cont) ou
aleatorios 0 ou 1 (bin)
#gera um novo vetor aleatório a cada loop

df <- data.frame(x,y,z) #junta x, y e z num dataframe
df <- as.geodata(df) #transforma a informação em geográfica
var <- variog(df, uvec=seq(0.044,0.44,by=0.044), tolerance=0.022)
#faz variograma considerando lag do gwr, omnidirecional

u <- var[[1]] #define vetores baseado na resposta do data.frame
v <- var[[2]]

matexp[i,] <- v

var.t <- variofit(var, cov.model = "pure.nugget", fix.nugget=F)

w <- var.t[[1]]

matteor[i,] <- w
}

write.table(matexp, "matexp_cfb2.txt", sep="\t") #escreve tabela de output com os vetores v
write.table(matteor, "matteor_cfb2.txt", sep="\t") #escreve tabela de output com os dados a
equação teórica

answer <- list(u,v) #output para conferência e de onde se pode retirar u, igual para todos os
vetores

```



```
return(answer)
}
```

Por exemplo, um retorno seriam os vetores:

```
u
0.032 0.064 0.096 0.128 0.160 0.192 0.224 0.256 0.288 0.320
v
0.2416821 0.2544959 0.2514554 0.2411800 0.2429262 0.2406688 0.2500626 0.2470828
0.2465506 0.2546178
```

Para esses dados, o efeito pepita puro corresponde ao valor de patamar (média) de 0,250, com desvio padrão de 1,461e-3.

Esse procedimento é realizado para 100 iterações e os resultados são analisados em conjunto de forma a confirmar a suposição de normalidade. Para isso utiliza-se o teste de *Kolmogorov-Smirnov* através da função existente no pacote básico do *R*. A função compara a distribuição analisada com a distribuição normal padrão.

```
> ks.test(matteor,"pnorm", mean(matteor$V1), sd(matteor$V1))
```

One-sample Kolmogorov-Smirnov test

```
data: matteor
D = 0.14277, p-value = 0.002
alternative hypothesis: two-sided
```

Considerando um nível de confiança de 95% rejeita-se a hipótese nula. Contudo, optou-se pela suposição de normalidade da distribuição, especialmente considerando o alto número de observações, de forma a dar prosseguimento à obtenção do indicador proposto.

8. Aplicação do Teste de Significância e demais análises

As demais etapas foram realizadas com o auxílio do *software Excel*.

Apêndice D – Exemplo de Cálculo de Indicador Local

Assim como realizado com o indicador Global, este item expõem o método de cálculo do indicador através do *software R* e os códigos desenvolvidos para isso.

No exemplo de cálculo são analisados os dados provenientes da Cidade Fictícia, apresentados no item 6.1. Algumas etapas são iguais as apresentadas no Apêndice C – Exemplo de Cálculo de Indicador Global, portanto apresentam comentários breves e referências ao item com o cálculo completo apresentado previamente.

1. Entrada dos dados

Os dados de entrada devem ser adicionados como explicado no respectivo item do cálculo do indicador global.

2. Regressão Linear

A regressão linear deve ser realizada como explicado no respectivo item do cálculo do indicador global.

3. Parâmetro de Suavização da Regressão Geograficamente Ponderada

O parâmetro de suavização é obtido como explicado no respectivo item do cálculo do indicador global.

4. Índice de Moran Local

O índice de Moran Local independe dos cálculos do indicador e pode ser realizado em qualquer momento do procedimento, desde que seja utilizado para comparação na etapa final. Considerando-se que a verificação de mais de 10% dos dados não possuírem vizinhos foi realizada para o índice de Moran Global deve-se seguir para o cálculo do índice Moran Local, ou LISA. Para isso foi utilizada a biblioteca *ncf* (BJORNSTAD, 2016).

O índice deve ser calculado para cada vizinhança. Para cada cálculo o retorno corresponde a um valor representativo da associação espacial e o p-valor correspondente para cada ponto.

```
> lisa.cd <- lisa(dados$X,dados$Y,dados$ESCOLHA,0.032)
```

Os resultados são apresentados na forma do indicador, p-valor, média dos vizinhos e número de pontos utilizados para o cálculo. Como o resultado é diferente ponto a ponto, a maneira mais eficiente de apresentação é através de mapas temáticos, o que pode ser construído através da função a seguir.

```
> plot.lisa(lisa.cd,inches = .03)
```

5. Semivariogramas experimentais

No cálculo do indicador global utiliza-se a biblioteca *geoR* (RIBEIRO; DIGGLE, 2016) para cálculo dos semivariogramas. Porém, não existem bibliotecas disponíveis no *software R* para semivariogramas locais. Dessa forma, elaborou-se um código para realizar este cálculo.

```
var.local <- function(x, y, z, lag, viz, resamp = 1000, quiet = FALSE){
```

```
  #análise ponto a ponto em uma matriz, a coluna é sempre o ponto analisado e as linhas seus vizinhos
```

```
  # dimensao do numero de atributos
```

```
  if (!is.null(dim(z))) {
```

```
    stop("\n z is multivariate.")
```

```
  }
```

```
  # iterador n vai receber o tamanho do vetor de atributos
```

```
  n <- length(z)
```

```
  #ajuste da distância de lag
```

```
  #####
```

```
  if (viz==1){ #apenas para a primeira vizinhança
```

```
    lag.min <- 0 #lag minimo
```

```

lag.max <- lag + tol.lag #lag maximo
} else if (viz==2){
lag.min <- lag - (tol.lag) #lag minimo
lag.max <- lag + (tol.lag) #lag maximo
}

#verifica se são dados contabilizáveis na equação
#####

dmat <- sqrt(outer(x, x, "-")^2 + outer(y, y, "-")^2) #cria matriz de distancias

#gera uma matriz de 1 e NA onde a conta eh valida pela distancia limite e angulo limite
dkl1 <- ifelse(dmat >= lag.min & dmat <= lag.max, 1, NA) #critério de distancias

dkl <- dkl1 #verifica se passa pelo critério da dist apenas OMNIDIRECIONAL

#cálculo do indicador
#####

negz = dkl * z #multiplica o mapa de distancias validas pelos valores de Z
smvar <- matrix(0, nrow = n, ncol = n)#cria mtx de semivariâncias

for (j in 1:n){ #loop para cálculo das semivariâncias em comparação com cada termo de z
(mesmo j de colunas)
  for (i in 1:n){
    smvar[i,j] <- (negz[i,j]-z[j])^2
  }
}

nlok <- apply(dkl, 2, sum, na.rm = TRUE) #realiza somatorio das colunas do mapa de dist dkl
= contar quantos termos de interesse temos

sum.smvar <- apply(smvar, 2, sum, na.rm = TRUE) #soma das semivariâncias por coluna =
por ponto
var <- rep(0,n)

```

```

var <- sum.smvar/(2*nlok) #cálculo da semivariância representativa de cada ponto

#teste de hipótese
#####

p <- NULL

if (resamp > 0) {
  perm <- matrix(NA, nrow = resamp, ncol = n) #estrutura para guardar os dados
  smvar2 <- matrix(0, nrow = n, ncol = n)#cria mtx de semivariâncias
  #####

  for (k in 1:resamp) { #a cada loop os dados são reorganizados e é calculado o valor de
semivariância

    #print interativo
    if (!quiet) {
      cat(k, " of ", resamp, "\n")
    }
    #####

    trekk <- sample(1:n) #sorteio aleatório do tamanho do item do vetor de z (ex tam 5: 5 2 3
1 4)
    z2 <- z[trekk]
    negz2= dkl * z2
    for (a in 1:n){ #loop para cálculo das semivariâncias em comparação com cada termo de z
(mesmo j de colunas)
      for (b in 1:n){
        smvar2[b,a] <- (negz2[b,a]-z2[a])^2
      }
    }
    sum.smvar2 <- apply(smvar2, 2, sum, na.rm = TRUE)
    perm[k,] <- sum.smvar2/(2*nlok)

```

```

    }
#####
    p <- (apply(var < t(perm), 1, sum))/(resamp + 1) #variancia e compara com cada linha do
perm, devolve a soma das linhas/n_perm
    p <- apply(cbind(p, 1 - p), 1, min) + 1/(resamp + 1) #verifica qual probabilidade eh menor
entre p e 1-p para todos os elementos e adota o menor (analise de probabilidades da curva
normal)
}

answer <- list(z =z, indicador = var, p = p, soma.semivar = sum.smvar, n = nlok, call =
deparse(match.call()))
answer

setwd("C:\\Users\\Claudia\\Dropbox\\Dissertacao\\Local_code")

id <- seq(1,n,1)

df <- data.frame(id,z,var,p,sum.smvar,nlok)

write.table(df,"result_ind_th.txt", sep="\t")
}

```

O resultado é similar ao do LISA, valores do indicador e p-valores associados para cada ponto dessa vizinhança. Como o objetivo é desenvolver um índice que permita a comparação entre pontos e não somente ponto a ponto, a significância obtida não foi utilizada. Para realizar o cálculo para as demais vizinhanças é necessário repetir todo o procedimento.

6. Semivariogramas teóricos

Para o cálculo do modelo teórico representativo de cada ponto foi utilizada uma modelagem de mínimos quadrados não lineares existente no pacote *minpack.lm* (ELZHOV *et al.*, 2016) do *software R*.

```
regres<- function(dados,lag){
```

```

#dados = colunas var_1,var_2...
#lag = vetor x nrow=10
n <- length(dados[,1]) #acha o tamanho n de linhas do BD
res.sph <- matrix(0, ncol=11, nrow=n) #matriz de modelos esféricos
res.exp <- matrix(0, ncol=11, nrow=n) #matriz de modelos exponenciais
res.gauss <- matrix(0, ncol=11, nrow=n) #matriz de modelos gaussianos
res.nugget <- matrix(0, ncol=11, nrow=n) #matriz de modelos exponenciais para os dados que
não convergiram em nenhum dos outros métodos
res <- matrix(0, ncol=11, nrow=n) #matriz de modelos finais

library("minpack.lm", lib.loc="~/R/win-library/3.4")
nls.control <- nls.lm.control(maxiter = 10000)

#foi feito loop a loop pois caso não convergisse (erro) não é executado o resto do código

for(i in 1:n){#loop esférico
  x <- lag
  y <- rep(0,10)
  tryCatch({
    for (j in 1:10) {
      y[j]<-dados[i,j]
    }
    for (k in 10:1){
      if (is.na(y[k])){
        y <- y[-k]
        x <- x[-k]
      }
    }
    model.sph <- 0
    model.sph <- nlsLM(y ~ ifelse( x < A, B + C*(1.5*x/A-0.5*(x/A)^3), B + C), start =
list(A=0.1, B= 0.1, C=0.1), algorithm="port", lower=c(0.001,0,0),control=nls.control)
    res.sph[i,1] <- summary(model.sph)$coefficients["A","Estimate"]
    res.sph[i,2] <- summary(model.sph)$coefficients["A","Std. Error"]
    res.sph[i,3] <- summary(model.sph)$coefficients["A","t value"]
  }, error=function(e){})
}

```

```

res.sph[i,4] <- summary(model.sph)$coefficients["B","Estimate"]
res.sph[i,5] <- summary(model.sph)$coefficients["B","Std. Error"]
res.sph[i,6] <- summary(model.sph)$coefficients["B","t value"]
res.sph[i,7] <- summary(model.sph)$coefficients["C","Estimate"]
res.sph[i,8] <- summary(model.sph)$coefficients["C","Std. Error"]
res.sph[i,9] <- summary(model.sph)$coefficients["C","t value"]
res.sph[i,10] <- cor(y,predict(model.sph))
res.sph[i,11] <- "sph"
}, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}

```

```

for(i in 1:n){ #loop exponencial
  x <- lag
  y <- rep(0,10)
  tryCatch({
    for (j in 1:10) {
      y[j]<-dados[i,j]
    }
    for (k in 10:1){
      if (is.na(y[k])){
        y <- y[-k]
        x <- x[-k]
      }
    }
    model.exp <- 0
    model.exp <- nlsLM((y ~ B + C*(1-exp(-x/A))), start = list(A=0.1, B= 0.1, C=0.1),
algorithm="port", lower=c(0.001,0,0),control=nlc)
    res.exp[i,1] <- summary(model.exp)$coefficients["A","Estimate"]
    res.exp[i,2] <- summary(model.exp)$coefficients["A","Std. Error"]
    res.exp[i,3] <- summary(model.exp)$coefficients["A","t value"]
    res.exp[i,4] <- summary(model.exp)$coefficients["B","Estimate"]
    res.exp[i,5] <- summary(model.exp)$coefficients["B","Std. Error"]
    res.exp[i,6] <- summary(model.exp)$coefficients["B","t value"]
    res.exp[i,7] <- summary(model.exp)$coefficients["C","Estimate"]

```



```

res.exp[i,8] <- summary(model.exp)$coefficients["C","Std. Error"]
res.exp[i,9] <- summary(model.exp)$coefficients["C","t value"]
res.exp[i,10] <- cor(y,predict(model.exp))
res.exp[i,11] <- "exp"
}, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}

for(i in 1:n){ #loop gaussiano
  x <- lag
  y <- rep(0,10)
  tryCatch({
    for (j in 1:10) {
      y[j]<-dados[i,j]
    }
    for (k in 10:1){
      if (is.na(y[k])){
        y <- y[-k]
        x <- x[-k]
      }
    }
    model.gauss <- 0
    model.gauss <- nlsLM((y ~ B + C*(1-exp(-3*(x/A)^2))), start = list(A=0.1, B= 0.1, C=0.1),
algorithm="port", lower=c(0.001,0,0),control=nlc)
    res.gauss[i,1] <- summary(model.gauss)$coefficients["A","Estimate"]
    res.gauss[i,2] <- summary(model.gauss)$coefficients["A","Std. Error"]
    res.gauss[i,3] <- summary(model.gauss)$coefficients["A","t value"]
    res.gauss[i,4] <- summary(model.gauss)$coefficients["B","Estimate"]
    res.gauss[i,5] <- summary(model.gauss)$coefficients["B","Std. Error"]
    res.gauss[i,6] <- summary(model.gauss)$coefficients["B","t value"]
    res.gauss[i,7] <- summary(model.gauss)$coefficients["C","Estimate"]
    res.gauss[i,8] <- summary(model.gauss)$coefficients["C","Std. Error"]
    res.gauss[i,9] <- summary(model.gauss)$coefficients["C","t value"]
    res.gauss[i,10] <- cor(y,predict(model.gauss))
    res.gauss[i,11] <- "gauss"

```

```

    }, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}

```

```

for(i in 1:n){ #loop gaussiano

```

```

  x <- lag

```

```

  y <- rep(0,10)

```

```

  tryCatch({

```

```

    for (j in 1:10) {

```

```

      y[j]<-dados[i,j]

```

```

    }

```

```

    for (k in 10:1){

```

```

      if (is.na(y[k])){

```

```

        y <- y[-k]

```

```

        x <- x[-k]

```

```

      }

```

```

    }

```

```

    model.nugget <- 0

```

```

    model.nugget <- mean(y)

```

```

    res.nugget[i,1] <- 0

```

```

    res.nugget[i,2] <- 0

```

```

    res.nugget[i,3] <- 0

```

```

    res.nugget[i,4] <- model.nugget

```

```

    res.nugget[i,5] <- 0

```

```

    res.nugget[i,6] <- 0

```

```

    res.nugget[i,7] <- 0

```

```

    res.nugget[i,8] <- 0

```

```

    res.nugget[i,9] <- 0

```

```

    res.nugget[i,10] <- 0

```

```

    res.nugget[i,11] <- "nugget"

```

```

  }, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}

```

```

for(i in 1:n) {

```

```

  if (res.sph[i,10] >= res.exp[i,10] & res.sph[i,10] >= res.gauss[i,10]){

```

```

    res[i,] <- res.sph[i,]
  }
  if (res.exp[i,10] >= res.sph[i,10] & res.exp[i,10] >= res.gauss[i,10]){
    res[i,] <- res.exp[i,]
  }
  if (res.gauss[i,10] >= res.sph[i,10] & res.gauss[i,10] >= res.exp[i,10]){
    res[i,] <- res.gauss[i,]
  }
  if (res.sph[i,10] == 0 & res.exp[i,10] == 0 & res.gauss[i,10] == 0){
    res[i,] <- res.nugget[i,]
  }
}

#print(res)
write.table(res.sph,"reg_sph_model.txt", sep="\t")
write.table(res.exp,"reg_exp_model.txt", sep="\t")
write.table(res.gauss,"reg_gauss_model.txt", sep="\t")
write.table(res.nugget,"reg_nugget_model.txt", sep="\t")
write.table(res,"reg_model.txt", sep="\t")
}

```

Assim, os arquivos exportados pelo código correspondem aos parâmetros calibrados de cada modelo e suas propriedades. É escolhido o modelo com maior correlação entre os dados da variável z e os dados previstos pelo modelo de z .

Os pontos que não apresentaram convergência com os modelos esférico, exponencial ou gaussiano são representados pelo efeito pepita puro.

7. Teste de Hipótese

O teste de hipótese é realizado de forma análoga ao teste de hipótese do indicador global, gerando diversos dados aleatórios, a partir dos quais é calculada a semivariância e modelados semivariogramas teóricos representativos. Os modelos são analisados em conjunto e comparados com uma distribuição normal.

```

teste7 <- function(x, y, lag, nloop=100, data.type){

  #análise ponto a ponto em uma matriz, a coluna é sempre o ponto analisado e as linhas seus
  vizinhos

  # iterador n vai receber o tamanho do vetor de atributos
  n <- length(x)

  #ajuste da distância de lag
  #####

  #cria uma matriz para resultados de cada lag
  vmat1 <- matrix(ncol=nloop, nrow=n)
  vmat2 <- matrix(ncol=nloop, nrow=n)
  vmat3 <- matrix(ncol=nloop, nrow=n)
  vmat4 <- matrix(ncol=nloop, nrow=n)
  vmat5 <- matrix(ncol=nloop, nrow=n)
  vmat6 <- matrix(ncol=nloop, nrow=n)
  vmat7 <- matrix(ncol=nloop, nrow=n)
  vmat8 <- matrix(ncol=nloop, nrow=n)
  vmat9 <- matrix(ncol=nloop, nrow=n)
  vmat10 <- matrix(ncol=nloop, nrow=n)

  dmat <- sqrt(outer(x, x, "-")^2 + outer(y, y, "-")^2) #cria matriz de distancias

  tol.lag<-lag/2

  #loop para gerar vários vetores aleatórios e testar seus resultados
  #####

  for (w in 1:nloop){

    if (data.type=='bin'){
      z <- runif(n,0,1)
    }
  }
}

```

```

for (m in 1:n){
  if (z[m]>=0.5){z[m]<- 1}
  else {z[m]<- 0}
}
}
if (data.type=='cont'){
  z <- runif(n,0,1)
}

for (q in 1:10) {

  if (q==1){

    lag.min <- 0 #lag minimo
    lag.max <- lag*q+tol.lag #lag maximo

  }
  else {

    lag.min <- lag*q-tol.lag #lag minimo
    lag.max <- lag*q+tol.lag #lag maximo

  }

  #gera uma matriz de 1 e NA onde a conta eh valida pela distancia limite e angulo limite
  dkl <- ifelse(dmat >= lag.min & dmat <= lag.max, 1, NA) #criterio de distancias
  #OMNIDIRECIONAL

  #cálculo do indicador
  #####

  negz = dkl * z #multiplica o mapa de distancias validas pelos valores de Z
  smvar <- matrix(0, nrow = n, ncol = n)#cria mtx de semivariâncias

```

```

    for (j in 1:n){ #loop para calculo das semivariâncias em comparação com cada termo de z
(mesmo j de colunas)
      for (i in 1:n){
        smvar[i,j] <- (negz[i,j]-z[j])^2
      }
    }

```

```

nlok <- apply(dkl, 2, sum, na.rm = TRUE) #realiza somatorio das colunas do mapa de dist
dkl = contar quantos termos de interesse temos

```

```

sum.smvar <- apply(smvar, 2, sum, na.rm = TRUE) #soma das semivariâncias por coluna
= por ponto

```

```

var <- rep(0,n)

```

```

var <- sum.smvar/(2*nlok) #calcula da semivariância representativa de cada ponto

```

```

if (q==1) {vmat1[,w] <- var} #atribui resultados para matriz do lag correspondente

```

```

if (q==2) {vmat2[,w] <- var}

```

```

if (q==3) {vmat3[,w] <- var}

```

```

if (q==4) {vmat4[,w] <- var}

```

```

if (q==5) {vmat5[,w] <- var}

```

```

if (q==6) {vmat6[,w] <- var}

```

```

if (q==7) {vmat7[,w] <- var}

```

```

if (q==8) {vmat8[,w] <- var}

```

```

if (q==9) {vmat9[,w] <- var}

```

```

if (q==10) {vmat10[,w] <- var}

```

```

rm(var)#apagar var

```

```

}

```

```

}

```

```

setwd("C:\\Users\\claudia.naizer\\Dropbox\\Dissertação\\Local_code")

```

```

write.table(vmat1,"var_viz1.txt", sep="\t")

```

```

write.table(vmat2,"var_viz2.txt", sep="\t")

```

```

write.table(vmat3,"var_viz3.txt", sep="\t")
write.table(vmat4,"var_viz4.txt", sep="\t")
write.table(vmat5,"var_viz5.txt", sep="\t")
write.table(vmat6,"var_viz6.txt", sep="\t")
write.table(vmat7,"var_viz7.txt", sep="\t")
write.table(vmat8,"var_viz8.txt", sep="\t")
write.table(vmat9,"var_viz9.txt", sep="\t")
write.table(vmat10,"var_viz10.txt", sep="\t")

}

```

Para facilitar a aplicação do índice optou-se por analisar os patamares médios das 100 simulações de cada ponto em conjunto, retornando pontos críticos iguais para o teste de significância de todos os itens.

```
> ks.test(dz,"pnorm",mean(dz$z),sd(dz$z))
```

One-sample Kolmogorov-Smirnov test

data: dz

D = 0.022725, p-value = 0.7382

alternative hypothesis: two-sided

Como pode-se observar os resultados são similares a distribuição normal para um nível de confiança de 95%.

8. Aplicação do Teste de Significância e demais análises

As demais etapas foram realizadas com o auxílio do *software Excel* e do *software QGIS*.

Apêndice E – Mapas correspondentes aos semivariogramas teóricos padronizados SIVAR-L

Primeiramente, são apresentados os mapas referentes ao indicador SIVAR-L e sua significância para 95% de confiança para as dez vizinhanças do banco de dados Cidade Fictícia.

Figura 9-6: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a segunda vizinhança resultante de análise local – Cidade Fictícia

Figura 9-7: Significância das Semivariâncias Teóricas Padronizadas para a segunda vizinhança resultante de análise local – Cidade Fictícia

Figura 9-8: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a terceira vizinhança resultante de análise local – Cidade Fictícia

Figura 9-9: Significância das Semivariâncias Teóricas Padronizadas para a terceira vizinhança resultante de análise local – Cidade Fictícia

Figura 9-10: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quarta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-11: Significância das Semivariâncias Teóricas Padronizadas para a quarta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-12: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quinta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-13: Significância das Semivariâncias Teóricas Padronizadas para a quinta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-14: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sexta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-15: Significância das Semivariâncias Teóricas Padronizadas para a sexta vizinhança resultante de análise local – Cidade Fictícia

Figura 9-16: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sétima vizinhança resultante de análise local – Cidade Fictícia

Figura 9-17: Significância das Semivariâncias Teóricas Padronizadas para a sétima vizinhança resultante de análise local – Cidade Fictícia

Figura 9-18: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a oitava vizinhança resultante de análise local – Cidade Fictícia

Figura 9-19: Significância das Semivariâncias Teóricas Padronizadas para a oitava vizinhança resultante de análise local – Cidade Fictícia

Figura 9-20: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a nona vizinhança resultante de análise local – Cidade Fictícia

Figura 9-21: Significância das Semivariâncias Teóricas Padronizadas para a nona vizinhança resultante de análise local – Cidade Fictícia

Figura 9-22: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a décima vizinhança resultante de análise local – Cidade Fictícia

Figura 9-23: Significância das Semivariâncias Teóricas Padronizadas para a décima vizinhança resultante de análise local – Cidade Fictícia

A seguir são apresentados os resultados do banco de dados do Centro de São Paulo – SP. Após a sexta vizinhança as observações significativas são praticamente as mesmas, dessa forma, os mapas seguintes não foram apresentados para não haver repetição.

Figura 9-24: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a segunda vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-25: Significância das Semivariâncias Teóricas Padronizadas para a segunda vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-26: : Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a terceira vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-27: Significância das Semivariâncias Teóricas Padronizadas para a terceira vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-28: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quarta vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-29: Significância das Semivariâncias Teóricas Padronizadas para a quarta vizinhança resultante de análise local – Centro, São Paulo – SP

Figura 9-30: : Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a quinta vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-31: Significância das Semivariâncias Teóricas Padronizadas para a quinta vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-32: Semivariâncias Teóricas Padronizadas, indicador SIVAR-L, para a sexta vizinhança resultante de análise local – Centro, São Paulo - SP

Figura 9-33: Significância das Semivariâncias Teóricas Padronizadas para a sexta vizinhança resultante de análise local – Centro, São Paulo – SP

Apêndice F – Mapas correspondentes ao índice LISA

Primeiramente, são apresentados os mapas referentes ao indicador LISA, correlação, significância para 95% de confiança e clusters para as dez vizinhanças do banco de dados Cidade Fictícia.

Figura 9-34: Correlação do Indicador LISA para a segunda vizinhança – Cidade Fictícia

Figura 9-35: Significância do Indicador LISA para a segunda vizinhança – Cidade Fictícia

Figura 9-36: Mapa de *Clusters* LISA para a segunda vizinhança – Cidade Fictícia

Figura 9-37: Correlação do Indicador LISA para a terceira vizinhança – Cidade Fictícia

Figura 9-38: Significância do Indicador LISA para a terceira vizinhança – Cidade Fictícia

Figura 9-39: Mapa de *Clusters* LISA para a terceira vizinhança – Cidade Fictícia

Figura 9-40: Correlação do Indicador LISA para a quarta vizinhança – Cidade Fictícia

Figura 9-41: Significância do Indicador LISA para a quarta vizinhança – Cidade Fictícia

Figura 9-42: Mapa de *Clusters* LISA para a quarta vizinhança – Cidade Fictícia

Figura 9-43: Correlação do Indicador LISA para a quinta vizinhança – Cidade Fictícia

Figura 9-44: Significância do Indicador LISA para a quinta vizinhança – Cidade Fictícia

Figura 9-45: Mapa de *Clusters* LISA para a quinta vizinhança – Cidade Fictícia

Figura 9-46: Correlação do Indicador LISA para a sexta vizinhança – Cidade Fictícia

Figura 9-47: Significância do Indicador LISA para a sexta vizinhança – Cidade Fictícia

Figura 9-48: Mapa de *Clusters* LISA para a sexta vizinhança – Cidade Fictícia

Figura 9-49: Correlação do Indicador LISA para a sétima vizinhança – Cidade Fictícia

Figura 9-50: Significância do Indicador LISA para a sétima vizinhança – Cidade Fictícia

Figura 9-51: Mapa de *Clusters* LISA para a sétima vizinhança – Cidade Fictícia

Figura 9-52: Correlação do Indicador LISA para a oitava vizinhança – Cidade Fictícia

Figura 9-53: Significância do Indicador LISA para a oitava vizinhança – Cidade Fictícia

Figura 9-54: Mapa de *Clusters* LISA para a oitava vizinhança – Cidade Fictícia

Figura 9-55: Correlação do Indicador LISA para a nona vizinhança – Cidade Fictícia

Figura 9-56: Significância do Indicador LISA para a nona vizinhança – Cidade Fictícia

Figura 9-57: Mapa de *Clusters* LISA para a nona vizinhança – Cidade Fictícia

Figura 9-58: Correlação do Indicador LISA para a décima vizinhança – Cidade Fictícia

Figura 9-59: Significância do Indicador LISA para a décima vizinhança – Cidade Fictícia

Figura 9-60: Mapa de *Clusters* LISA para a décima vizinhança – Cidade Fictícia

A seguir são apresentados os resultados do banco de dados do Centro de São Paulo – SP. Após a sétima vizinhança as observações significativas são similares e a correlação LISA apresenta valores baixos, dessa forma, os mapas seguintes não foram apresentados para não haver repetição.

Figura 9-61: Correlação do Indicador LISA para a segunda vizinhança – Centro, São Paulo – São Paulo

Figura 9-62: Significância do Indicador LISA para a segunda vizinhança – Centro, São Paulo – São Paulo

Figura 9-63: Mapa de *Clusters* LISA para a segunda vizinhança – Centro, São Paulo – São Paulo

Figura 9-64: Correlação do Indicador LISA para a terceira vizinhança – Centro, São Paulo – São Paulo

Figura 9-65: Significância do Indicador LISA para a terceira vizinhança – Centro, São Paulo – São Paulo

Figura 9-66: Mapa de *Clusters* LISA para a terceira vizinhança – Centro, São Paulo – São Paulo

Figura 9-67: Correlação do Indicador LISA para a quarta vizinhança – Centro, São Paulo – São Paulo

Figura 9-68: Significância do Indicador LISA para a quarta vizinhança – Centro, São Paulo – São Paulo

Figura 9-69: Mapa de *Clusters* LISA para a quarta vizinhança – Centro, São Paulo – São Paulo

Figura 9-70: Correlação do Indicador LISA para a quinta vizinhança – Centro, São Paulo – São Paulo

Figura 9-71: Significância do Indicador LISA para a quinta vizinhança – Centro, São Paulo – São Paulo

Figura 9-72: Mapa de *Clusters* LISA para a quinta vizinhança – Centro, São Paulo – São Paulo

Figura 9-73: Correlação do Indicador LISA para a sexta vizinhança – Centro, São Paulo – São Paulo

Figura 9-74: Significância do Indicador LISA para a sexta vizinhança – Centro, São Paulo – São Paulo

Figura 9-75: Mapa de *Clusters* LISA para a sexta vizinhança – Centro, São Paulo – São Paulo

Figura 9-76: Correlação do Indicador LISA para a sétima vizinhança – Centro, São Paulo – São Paulo

Figura 9-77: Significância do Indicador LISA para a sétima vizinhança – Centro, São Paulo – São Paulo

Figura 9-78: Mapa de *Clusters* LISA para a sétima vizinhança – Centro, São Paulo – São Paulo