



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CTC - CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA  
DE TRANSPORTES E GESTÃO TERRITORIAL

CARLOS AUGUSTO ZILLI

**REGRESSÃO GEOGRAFICAMENTE PONDERADA APLICADA  
NA AVALIAÇÃO EM MASSA DE IMÓVEIS URBANOS**

FLORIANÓPOLIS  
2020

CARLOS AUGUSTO ZILLI

**REGRESSÃO GEOGRAFICAMENTE PONDERADA APLICADA  
NA AVALIAÇÃO EM MASSA DE IMÓVEIS URBANOS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia de Transportes e Gestão Territorial.

Orientador Prof. Dr. Norberto Hochheim

Florianópolis

2020

Ficha de identificação da obra elaborada pelo autor, 2020, através do  
Programa de Geração Automática da Biblioteca Universitária da UFSC.

Zilli, Carlos Augusto

Regressão geograficamente ponderada aplicada na avaliação  
em massa de imóveis urbanos / Carlos Augusto Zilli;  
Orientador, Norberto Hochheim, 2020.

192 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro Tecnológico, Programa de Pós-Graduação em  
Engenharia de Transportes e Gestão Territorial,  
Florianópolis, 2020.

Inclui referências.

1. Engenharia de Transportes e Gestão Territorial. 2.  
Regressão Geograficamente Ponderada. 3. Avaliação em Massa  
de Imóveis. 4. Avaliação de Imóveis. 5. Planta de Valores  
Genéricos. I. Hochheim, Norberto. II. Universidade Federal  
de Santa Catarina. Programa de Pós-Graduação em Engenharia  
de Transportes e Gestão Territorial. III. Título.

CARLOS AUGUSTO ZILLI

**REGRESSÃO GEOGRAFICAMENTE PONDERADA APLICADA  
NA AVALIAÇÃO EM MASSA DE IMÓVEIS URBANOS**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

**Prof. Everton da Silva, Dr.**

Universidade Federal de Santa Catarina – UFSC

**Profa. Marlene Salete Uberti, Dra.**

Universidade Federal Rural do Rio de Janeiro – UFRJ

**Prof. Rubens Alves Dantas, Dr.**

Universidade Federal de Pernambuco – UFPE

**Profa. Vera do Carmo Comparsi de Vargas, Dra.**

Universidade Federal de Santa Catarina – UFSC

Certificamos que esta é a **VERSAO ORIGINAL E FINAL** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia de Transportes e Gestão Territorial.

---

**Prof. Dra. Ana Maria Bencciveni Franzoni**

Coordenadora do Programa

---

**Prof. Dr. Norberto Hochheim**

Orientador e Moderador

Florianópolis, 27 de julho de 2020

## **DEDICATÓRIA**

*In memoriam*, dedico este trabalho a ti, pai, que acometido por alzheimer, esqueceu de muitos, mas não de mim. Esse trabalho é nosso. Obrigado!

## **AGRADECIMENTOS**

Inicialmente agradeço a Deus pelo dom da vida e por permitir com que eu pudesse concluir esse curso de mestrado em uma universidade que conheci durante o curso e pela qual eu me tornei completamente apaixonado. Obrigado Deus!

E, desta forma, estendo os meus agradecimentos à (ao):

- minha mãe, Zoleide, que sempre esteve ao meu lado, me apoiando, independente do que eu escolhesse. Mãe, tu és um exemplo de otimismo e perseverança!
- meu amigo, Murilo, irmão da vida, atual sócio, que topou fazer esse mestrado comigo e me acompanhou nessa loucura de ir semanalmente para a capital;
- meu amigo, Droubi, pelas inúmeras discussões e ajudas nesse estudo e por ter um vulcão de ideias em sua cabeça e sempre me chamar para fazer parte delas;
- minhas amigas, Daiane e Fabiane, por me incentivarem, auxiliarem e às vezes até me acompanharem até Florianópolis. Daiane, obrigado pelas contribuições;
- meu orientador, professor Norberto, por me ajudar, aconselhar e acreditar que poderíamos, juntos, fazer um trabalho que em algum momento teria um fim;
- banca examinadora, professores Everton, Vera, Marlene e Rubens, por aceitarem meu convite e por contribuírem para a melhoria desse trabalho de mestrado.

*“Descobri como é bom chegar quando se tem paciência. E para se chegar, onde quer que seja, aprendi que não é preciso dominar a força, mas a razão. É preciso, antes de qualquer coisa, querer”.*

*– AMIR KLINK*

## RESUMO

O valor venal de um imóvel para fins tributários, em geral, é determinado por meio de processos de avaliação em massa. Nesse processo, um desafio comum é encontrar um modelo que seja capaz de tratar, mantendo razoável acurácia, os efeitos decorrentes da dependência e da heterogeneidade espacial. Este estudo objetivou investigar o uso de Regressão Geograficamente Ponderada (GWR) na avaliação em massa de imóveis urbanos de modo a reduzir esses efeitos. Para tanto, essa investigação se deu em imóveis tipo apartamento nos bairros Centro, Agronômica e Trindade, no município de Florianópolis/SC, contando, para isso, com uma amostra de trabalho contendo 190 dados e uma amostra de validação contendo 35 dados de apartamentos. A proposta metodológica consistiu em investigar o uso do modelo GWR e comparar suas estatísticas com as obtidas pela regressão clássica (OLS) e espacial (SAR/CAR). Os resultados mostraram a eficácia da modelagem GWR em reduzir os efeitos espaciais presentes nos dados da amostra. O modelo GWR mostrou-se mais aderente e superior às modelagens anteriores em todos os indicadores de qualidade (AIC, BIC, LIK,  $R^2$  e RMSE) e desempenho (Med R, COD e PRD), com valores dentro dos limites recomendados pela IAAO. Os métodos de regressão abordados neste estudo foram, também, utilizados de forma combinada com o interpolador de krigagem ordinária para estimar o valor da localização dos imóveis e gerar uma superfície de valores unitários homogeneizados representativa dos bairros em estudo. Para krigagem das superfícies de valores unitários, utilizou-se o semivariograma exponencial, com o qual se constatou que a superfície gerada pelos valores preditos pelo modelo GWR apresentou menor valor para o RMSE e com indicadores de desempenho dentro dos limites definidos pela IAAO. O método proposto permitiu avaliar a área em estudo e constatar que o uso da modelagem GWR pode ser utilizada na avaliação em massa de imóveis urbanos.

**Palavras-Chave:** Regressão Geograficamente Ponderada. Avaliação de Imóveis. Planta de Valores Genéricos. Economia Imobiliária. Engenharia de Avaliações.

## ABSTRACT

The market value of a property for tax purposes, in general, is determined through mass appraisal processes. In this process, a common challenge is to find a model that is capable of treating, while maintaining reasonable accuracy, the effects of dependence and spatial heterogeneity. Thus, in order to minimize these effects, this study aimed to investigate the use of Geographically Weighted Regression (GWR) in the mass evaluation of urban properties. To this end, this investigation took place in apartment-type properties in the Centro, Agronômica and Trindade neighborhoods, in the city of Florianópolis/SC, counting, for this purpose, with a work sample containing 190 data from apartments and a validation sample containing 35 data. The methodological proposal consisted of investigating the use of the GWR model and comparing its statistics with those obtained by classical (OLS) and spatial (SAR/CAR) regression. The results showed the effectiveness of GWR modeling in reducing the spatial effects present in the sample data. The GWR model proved to be more adherent and superior to the previous models in all quality (AIC, BIC, LIK,  $R^2$  and RMSE) and performance (Med R, COD and PRD) indicators, with values within the limits recommended by the IAAO. The regression methods addressed in this study were also used in combination with the ordinary kriging interpolator to estimate the value of the location of the properties and to generate a homogenized unit value surface representative of the neighborhoods under study. For kriging the surfaces of unit values, the exponential semivariogram was used, with which it was found that the surface generated by the values predicted by the GWR model presented a lower value for the RMSE and with performance indicators within the limits defined by the IAAO. The proposed method allowed us to evaluate the area under study and to verify that the use of GWR modeling can be used in the mass appraisal of urban properties.

**Keywords:** Geographically Weighted Regression. Real Estate Appraisal. Generic Value Plan. Real Estate Economics. Appraisal Engineering.

## SINTESI

In genere il valore di mercato di un immobile, ai fini fiscali, viene definito attraverso processi di valutazione di massa. In questo processo, solitamente la difficoltà sta nel reperire un modello che, pur mantenendo un'accuratezza ragionevole, consideri gli effetti della dipendenza e dell'eterogeneità spaziale. Pertanto, al fine di minimizzare questi effetti, l'obiettivo di questa analisi era studiare l'uso della GWR (regressione geograficamente ponderata) nella valutazione di massa delle proprietà urbane. A tal fine, la presente indagine è stata svolta su appartamenti ubicati nei quartieri Centro, Agronômica e Trindade della città di Florianópolis/SC, contando, a tal fine, con un campione di lavoro contenente 190 dati provenienti da appartamenti ed un campione di validazione contenente 35 dati. La proposta metodologica consisteva nello studio dell'uso del modello GWR e nel confronto delle statistiche risultanti con quelle ottenute dalla regressione lineare (OLS) e spaziale (SAR/CAR). I risultati hanno mostrato l'efficacia del modello GWR nella riduzione degli effetti spaziali presenti nei dati del campione. Il modello GWR si è dimostrato più aderente e superiore ai modelli precedenti in tutti gli indicatori di qualità (AIC, BIC, LIK, R<sup>2</sup> e RMSE) e di prestazioni (Med R, COD e PRD), con valori entro i limiti raccomandati dall'IAAO. I metodi di regressione applicati in questo studio sono stati anche usati in combinazione con la comune interpolazione kriging, per stimare il valore della posizione degli immobili e per generare un valore unitario omogeneo della superficie, rappresentativo dei quartieri studiati. Per il kriging delle superfici dei valori unitari è stato utilizzato il semivariogramma esponenziale, con il quale è stato riscontrato che la superficie generata dai valori previsti dal modello GWR presentava un valore inferiore per l'RMSE e con indicatori di prestazione entro i limiti definiti dall'IAAO. Il metodo proposto ci ha permesso di valutare l'area in esame e di verificare che l'uso del modello GWR può essere utilizzato nella valutazione di massa delle proprietà urbane.

**Parole Chiave:** Regressione geograficamente ponderata. Valutazione Immobiliare. Piano di valore generico. Economia immobiliare. Ingegneria della valutazione.

## LISTA DE FIGURAS

Figura 1: Recorte de uma PVG com o valor unitário por face de quadra .....	36
Figura 2: Fluxograma para elaboração de uma planta de valores .....	37
Figura 3: Fluxograma de decisão sobre o modelo adequado .....	55
Figura 4: Noções de Contiguidade de Rainha (A) e Torre (B).....	57
Figura 5: Esquema de <i>Kernel Espacial</i> e <i>Bandwidth</i> .....	66
Figura 6: Funções de Ponderação Espacial com <i>Bandwidth</i> Fixo.....	69
Figura 7: Funções de Ponderação Espacial com <i>Bandwidth</i> Variável .....	69
Figura 8: Esquema contendo os modelos analisados .....	82
Figura 9: Ponte Hercílio Luz (Florianópolis) .....	87
Figura 10: Figura Representativa de Florianópolis - Santa Catarina.....	88
Figura 11: Localização dos Núcleos Distritais de Florianópolis.....	90
Figura 12: Bairros Estudados em Florianópolis.....	92
Figura 13: Distribuição espacial dos dados de mercado da amostra de trabalho .....	94
Figura 14: Distribuição espacial dos dados da amostra de validação .....	95
Figura 15: Histograma de frequências de VU e de ln (VU) .....	100
Figura 16: Histograma da variável transformada ln (VU) .....	101
Figura 17: Gráfico Box-Cox do valor unitário não transformado.....	102
Figura 18: Gráfico <i>boxplot</i> com a variável VU e ln(VU) .....	103
Figura 19: Diagrama de cores indicando as correlações entre as variáveis .....	105
Figura 20: Gráfico de dispersão e <i>boxplot</i> da variável área privativa (AP).....	107
Figura 21 A e B: Gráfico <i>boxplot</i> das variáveis dormitórios (ND) e banheiros (NB)	108
Figura 22 A e B: Gráfico <i>boxplot</i> das variáveis vagas (NG) e mobília (MO).....	109
Figura 23 A e B: Gráfico <i>boxplot</i> das variáveis piscina (PS) e padrão (PC).....	110
Figura 24: Gráfico <i>boxplot</i> representando os bairros em estudo.....	110
Figura 25: Distribuição espacial dos resíduos do modelo OLS sem bairro .....	112
Figura 26: Distribuição espacial dos resíduos do modelo OLS com bairro .....	113
Figura 27: Gráfico de resíduos x valores estimados pelo modelo .....	118
Figura 28 A e B: Gráfico <i>boxplot</i> e histograma de resíduos padronizados.....	119
Figura 29: Gráfico de elementos Hii x quadrado dos resíduos padronizados .....	121
Figura 30: Gráfico de resíduos padronizados x ln (VU) estimado .....	122
Figura 31: Diagrama de espalhamento de Moran para a variável ln (VU).....	125

Figura 32: Semivariograma experimental de ln do valor unitário observado .....	126
Figura 33: Curva gaussiana fixa para a <i>bandwidth</i> hipotética .....	134
Figura 34: Diagrama de dispersão para modelo OLS <sub>CB</sub> .....	141
Figura 35: Diagrama de dispersão para modelo CAR <sub>SB</sub> .....	141
Figura 36: Diagrama de dispersão para modelo GWR <sub>SB</sub> .....	142
Figura 37: Superfície de erros relativos absolutos para a modelagem OLS <sub>CB</sub> .....	143
Figura 38: Superfície de erros relativos absolutos para a modelagem CAR <sub>SB</sub> .....	143
Figura 39: Superfície de erros relativos absolutos para a modelagem GWR <sub>SB</sub> .....	144
Figura 40 A a D: Erros relativos da amostra de validação para as modelagens .....	147
Figura 41 A a C: Mapas variográficos dos efeitos direcionais dos modelos.....	149
Figura 42: Semivariograma experimental dos valores unitários .....	151
Figura 43: Semivariograma teórico exponencial para regressão OLS <sub>CB</sub> .....	153
Figura 44: Superfície de predição dos valores unitários para OLS <sub>CB</sub> .....	154
Figura 45: Superfície predição do erro para OLS <sub>CB</sub> .....	154
Figura 46: Semivariograma experimental dos valores unitários .....	155
Figura 47: Semivariograma teórico exponencial para regressão CAR <sub>SB</sub> .....	157
Figura 48: Superfície de predição dos valores unitários para CAR <sub>SB</sub> .....	157
Figura 49: Superfície predição do erro para CAR <sub>SB</sub> .....	158
Figura 50: Semivariograma experimental dos valores unitários .....	159
Figura 51: Semivariograma teórico exponencial para regressão GWR <sub>SB</sub> .....	160
Figura 52: Superfície de predição dos valores unitários para GWR <sub>SB</sub> .....	161
Figura 53: Superfície predição do erro para GWR <sub>SB</sub> .....	162
Figura 54: Comparação entre a krigagem das modelagens para o Centro .....	163
Figura 55: Comparação entre a krigagem das modelagens para a Trindade .....	163
Figura 56: Comparação entre a krigagem das modelagens para a Agronômica....	164

## LISTA DE TABELAS

Tabela 1: Características dos Modelos Globais e Modelos Locais .....	60
Tabela 2: Funções de Ponderação ou <i>Kernels</i> .....	68
Tabela 3: Simulação de valores de pesos $w_{ij}$ .....	70
Tabela 4: Quantidade de amostra de trabalho e de validação em cada bairro .....	94
Tabela 5: Medidas de tendência central da variável explicada .....	100
Tabela 6: Coeficientes de assimetria e curtose da variável explicada .....	101
Tabela 7: Estatísticas descritivas da variável explicada VU e In (VU).....	103
Tabela 8: Matriz de correlações das variáveis explicativas .....	105
Tabela 9: Significâncias das variáveis excluídas da fase de modelagem .....	106
Tabela 10: Estatísticas relativas aos parâmetros do modelo OLS <sub>SB</sub> .....	112
Tabela 11: Estatísticas relativas aos parâmetros do modelo OLS <sub>CB</sub> .....	115
Tabela 12: Estatísticas relativas ao modelo de regressão OLS <sub>CB</sub> .....	116
Tabela 13: Análise da variância do modelo de regressão OLS <sub>CB</sub> .....	117
Tabela 14: Distribuição dos resíduos normalizados (OLS <sub>CB</sub> ) .....	119
Tabela 15: Matriz de correlações das variáveis transformadas.....	120
Tabela 16: Teste do Fator de Inflação da Variância (VIF).....	120
Tabela 17: Teste global de Moran para matrizes de ponderação espacial .....	124
Tabela 18: Teste dos Multiplicadores de Lagrange ( <i>LM</i> ) com e sem bairro .....	127
Tabela 19: Estatísticas dos modelos do erro e da defasagem espacial.....	128
Tabela 20: Estatísticas relativas aos parâmetros do modelo CAR <sub>SB</sub> .....	130
Tabela 21: Estatísticas relativas ao modelo de regressão CAR <sub>SB</sub> .....	130
Tabela 22: Escolha pelo Critério de Informação de Akaike (A/C).....	133
Tabela 23: Escolha pelo Critério de Akaike Corrigido (A/Cc) .....	133
Tabela 24: Escolha pelo Critério de Validação Cruzada (CV) .....	133
Tabela 25: Teste de estacionariedade dos parâmetros do modelo GWR <sub>SB</sub> .....	135
Tabela 26: Estatística relativa à variável PS do modelo GWR <sub>SB</sub> .....	136
Tabela 27: Estatísticas relativas aos parâmetros do modelo GWR <sub>SB</sub> .....	137
Tabela 28: Estatísticas relativas ao modelo de regressão GWR <sub>SB</sub> .....	138
Tabela 29: Indicadores de avaliação da qualidade dos modelos de regressão .....	139
Tabela 30: Raiz quadrada do erro quadrático médio da amostra de trabalho.....	142
Tabela 31: Limites recomendados para os parâmetros de avaliação em massa....	145

Tabela 32: Valores calculados para os indicadores de avaliação em massa.....	146
Tabela 33: Raiz quadrada do erro quadrático médio da amostra de validação .....	146
Tabela 34: Valor do erro relativo absoluto médio das amostras de validação .....	148
Tabela 35: Apartamento paradigma para homeogeneização.....	150
Tabela 36: Parâmetros dos semivariogramas teóricos do ajuste OLS <sub>CB</sub> .....	152
Tabela 37: Validação cruzada entre os modelos teóricos e o ajuste OLS <sub>CB</sub> .....	152
Tabela 38: Parâmetros dos semivariogramas teóricos do ajuste CAR <sub>SB</sub> .....	156
Tabela 39: Validação cruzada entre os modelos teóricos e o ajuste CAR <sub>SB</sub> .....	156
Tabela 40: Parâmetros dos semivariogramas teóricos do ajuste GWR <sub>SB</sub> .....	159
Tabela 41: Validação cruzada entre os modelos teóricos e o ajuste GWR <sub>SB</sub> .....	160
Tabela 42: Indicadores de qualidade da krigagem ordinária.....	164
Tabela 43: Valor do RMSE para krigagem dos valores preditos pelos modelos .....	165
Tabela 44: Limites recomendados para os parâmetros de avaliação em massa....	166
Tabela 45: Valores calculados para os parâmetros de avaliação em massa.....	166

## LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas

AIC Critério de Informação de Akaike

BIC Critério de Informação Bayesiano

CAR Regressão do Erro Espacial

COD Coeficiente de Dispersão da Mediana

CTM Cadastro Territorial Multifinalitário

CTN Código Tributário Nacional

CV Validação Cruzada

GWR Regressão Geograficamente Ponderada

IAAO *International Association of Assessing Officers*

IPTU Imposto Predial e Territorial Urbano

ITBI Imposto sobre Tramitação de Bens Imóveis

LIK Função de Verossimilhança

LM Multiplicador de Lagrange

Med R Mediana das Razões de Avaliação

MSE Erro Quadrático Médio

MWR Regressão de Janelas Móveis

NBR Norma Brasileira

OLS Regressão dos Mínimos Quadrados Ordinários

PRD Diferencial Relativo ao Preço

PVG Planta de Valores Genéricos

RGP Regressão Geograficamente Ponderada

RMS Raiz do Valor Quadrático Médio

RMSE Raiz Quadrada do Erro Quadrático Médio

SAR Regressão da Defasagem Espacial

UTM Universal Transversa de Mercator

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>19</b>
1.1	OBJETIVOS .....	21
1.1.1	<b>Objetivo Geral .....</b>	<b>21</b>
1.1.2	<b>Objetivos Específicos .....</b>	<b>21</b>
1.2	DELIMITAÇÃO DA PESQUISA .....	22
1.3	JUSTIFICATIVA.....	22
1.4	ESTADO DE ARTE DA GWR NA AVALIAÇÃO DE IMÓVEIS.....	24
1.5	ESTRUTURA DA PESQUISA .....	28
<b>2</b>	<b>ENGENHARIA DE AVALIAÇÕES DE IMÓVEIS .....</b>	<b>29</b>
2.1	AVALIAÇÃO EM MASSA DE IMÓVEIS.....	31
2.1.1	<b>Conceitos e Métodos de Avaliação em Massa.....</b>	<b>31</b>
2.1.2	<b>Análise de Desempenho da Avaliação em Massa .....</b>	<b>33</b>
2.2	PLANTA DE VALORES GENÉRICOS (PVG) .....	35
2.2.1	<b>Cadastro Territorial Multifinalitário (CTM).....</b>	<b>37</b>
2.2.2	<b>Imposto Predial e Territorial Urbano (IPTU).....</b>	<b>38</b>
<b>3</b>	<b>MODELOS DE REGRESSÃO CLÁSSICA E ESPACIAL .....</b>	<b>40</b>
3.1	MODELO DE REGRESSÃO CLÁSSICA.....	40
3.1.1	<b>Inferência Estatística em Regressões Lineares.....</b>	<b>43</b>
3.2	MODELOS DE REGRESSÃO ESPACIAL.....	47
3.2.1	<b>Modelo da Defasagem Espacial (<i>Spatial Lag</i>) .....</b>	<b>50</b>
3.2.2	<b>Modelo do Erro Espacial (<i>Spatial Error</i>) .....</b>	<b>51</b>
3.2.3	<b>Estimação do Modelo de Regressão Espacial .....</b>	<b>52</b>
3.2.4	<b>Testes de Especificação .....</b>	<b>52</b>
3.2.4.1	Teste I de Índice Global de Moran.....	53
3.2.4.2	Teste dos Multiplicadores de Lagrange .....	54
3.2.5	<b>Matrizes de Ponderação Espacial ou Peso Espacial.....</b>	<b>56</b>
<b>4</b>	<b>REGRESSÃO GEOGRAFICAMENTE PONDERADA .....</b>	<b>59</b>
4.1	MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA.....	60
4.2	PONDERAÇÃO ESPACIAL E PROCESSO DE CALIBRAÇÃO .....	64
4.2.1	<b>Funções de Ponderação Espacial.....</b>	<b>64</b>
4.2.2	<b>Processo de Calibração da Função de Ponderação Espacial .....</b>	<b>71</b>
4.3	TESTES APLICADOS À REGRESSÃO GEOPONDERADA.....	73

<b>4.3.1</b>	<b>Teste de Estacionaridade dos Parâmetros .....</b>	<b>74</b>
<b>4.3.2</b>	<b>Teste de Autocorrelação dos Resíduos .....</b>	<b>75</b>
<b>4.3.3</b>	<b>Identificação dos Coeficientes Constantes do Modelo.....</b>	<b>76</b>
<b>4.4</b>	<b>MEDIDAS DE QUALIDADE DE AJUSTE DO MODELO .....</b>	<b>77</b>
<b>5</b>	<b>MATERIAIS, METODOLOGIA E ÁREA DE ESTUDO.....</b>	<b>80</b>
<b>5.1</b>	<b>CLASSIFICAÇÃO DA PESQUISA.....</b>	<b>80</b>
<b>5.2</b>	<b>PROCEDIMENTOS METODOLÓGICOS .....</b>	<b>82</b>
<b>5.3</b>	<b>CARACTERIZAÇÃO DA ÁREA DE ESTUDO .....</b>	<b>86</b>
<b>5.3.1</b>	<b>Município de Florianópolis: Capital de Santa Catarina .....</b>	<b>86</b>
<b>5.3.2</b>	<b>Breve Histórico de Florianópolis.....</b>	<b>88</b>
<b>5.3.3</b>	<b>Geografia de Florianópolis .....</b>	<b>89</b>
<b>5.3.4</b>	<b>Aspectos Sócio-Econômicos .....</b>	<b>92</b>
<b>5.4</b>	<b>COLETA DE DADOS DO MERCADO IMOBILIÁRIO .....</b>	<b>93</b>
<b>5.4.1</b>	<b>Descrição das Variáveis da Amostra .....</b>	<b>95</b>
<b>6</b>	<b>RESULTADOS E ANÁLISES .....</b>	<b>99</b>
<b>6.1</b>	<b>ANÁLISE EXPLORATÓRIA DOS DADOS DA AMOSTRA.....</b>	<b>99</b>
<b>6.1.1</b>	<b>Análise da Variável Explicada (Valor Unitário).....</b>	<b>99</b>
<b>6.1.2</b>	<b>Análise das Variáveis Explicativas .....</b>	<b>104</b>
<b>6.2</b>	<b>MODELO CLÁSSICO DE REGRESSÃO LINEAR MÚLTIPLA .....</b>	<b>111</b>
<b>6.3</b>	<b>MODELO CLÁSSICO COM VARIÁVEIS DE LOCALIZAÇÃO (OLS<sub>CB</sub>) .....</b>	<b>114</b>
<b>6.3.1</b>	<b>Estatísticas dos Regressores do Modelo.....</b>	<b>114</b>
<b>6.3.2</b>	<b>Estatísticas do Modelo de Regressão .....</b>	<b>115</b>
<b>6.3.3</b>	<b>Pressupostos Básicos do Modelo de Regressão .....</b>	<b>117</b>
<b>6.4</b>	<b>MODELO DE REGRESSÃO ESPACIAL .....</b>	<b>123</b>
<b>6.4.1</b>	<b>Estimação da Matriz de Pesos Espaciais .....</b>	<b>123</b>
<b>6.4.2</b>	<b>Indicador de Autocorrelação Espacial: Multiplicador de Lagrange .....</b>	<b>127</b>
<b>6.4.3</b>	<b>Estimação do Modelo de Regressão Espacial do Erro (CAR) .....</b>	<b>129</b>
<b>6.4.4</b>	<b>Estatísticas dos Regressores do Modelo .....</b>	<b>129</b>
<b>6.4.5</b>	<b>Estatísticas do Modelo de Regressão .....</b>	<b>130</b>
<b>6.5</b>	<b>MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA.....</b>	<b>132</b>
<b>6.5.1</b>	<b>Teste DIFF de Estacionaridade dos Parâmetros do Modelo.....</b>	<b>134</b>
<b>6.5.2</b>	<b>Identificação de Coeficientes Constantes no Modelo .....</b>	<b>136</b>
<b>6.5.3</b>	<b>Modelo Misto de Regressão Geograficamente Ponderada.....</b>	<b>136</b>

<b>6.5.4</b>	<b>Estatísticas dos Regressores do Modelo .....</b>	<b>137</b>
<b>6.5.5</b>	<b>Estatísticas do Modelo de Regressão .....</b>	<b>138</b>
<b>6.6</b>	<b>PERFORMANCE DOS MODELOS DE REGRESSÃO.....</b>	<b>139</b>
<b>6.6.1</b>	<b>Qualidade do Ajuste dos Modelos .....</b>	<b>139</b>
<b>6.6.2</b>	<b>Desempenho dos Modelos de Regressão.....</b>	<b>144</b>
<b>6.7</b>	<b>SUPERFÍCIES DE VALORES UNITÁRIOS: KRIGAGEM ORDINÁRIA ....</b>	<b>149</b>
<b>6.7.1</b>	<b>Considerações Iniciais: Superfícies de Valores .....</b>	<b>150</b>
<b>6.7.2</b>	<b>Superfície de Predição da Regressão Linear Clássica .....</b>	<b>151</b>
6.7.2.1	Modelagem dos Semivariogramas OLS <sub>CB</sub> .....	151
6.7.2.2	Superfície de Valores Preditos pelo Modelo OLS <sub>CB</sub> .....	153
<b>6.7.3</b>	<b>Superfície de Predição da Regressão do Erro Espacial .....</b>	<b>155</b>
6.7.3.1	Modelagem dos Semivariogramas CAR <sub>SB</sub> .....	155
6.7.3.2	Superfície de Valores Preditos pelo Modelo CAR <sub>SB</sub> .....	157
<b>6.7.4</b>	<b>Superfície de Predição da Regressão Geograficamente Ponderada ...</b>	<b>158</b>
6.7.4.1	Modelagem dos Semivariogramas GWR <sub>SB</sub> .....	159
6.7.4.2	Superfície de Valores Preditos pelo Modelo GWR <sub>SB</sub> .....	161
<b>6.7.5</b>	<b>Indicadores de Desempenho da Krigagem .....</b>	<b>164</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>168</b>
7.1	CONCLUSÕES.....	168
<b>7.1.1</b>	<b>Variáveis do modelo clássico de regressão linear múltipla (OLS<sub>CB</sub>) ...</b>	<b>168</b>
<b>7.1.2</b>	<b>Matriz de pesos e autocorrelação na regressão espacial (CAR<sub>SB</sub>) .....</b>	<b>168</b>
<b>7.1.3</b>	<b>Performance da regressão geograficamente ponderada (GWR<sub>SB</sub>) .....</b>	<b>169</b>
<b>7.1.4</b>	<b>Interpolação dos valores unitários por krigagem ordinária .....</b>	<b>170</b>
<b>7.1.5</b>	<b>Avaliação em massa de apartamentos da região estudada .....</b>	<b>171</b>
7.2	RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	172
<b>REFERÊNCIAS.....</b>		<b>173</b>
APÊNDICE 01: AMOSTRA DE DADOS DE MERCADO (TRABALHO).....		181
APÊNDICE 02: AMOSTRA DE DADOS DE MERCADO (VALIDAÇÃO).....		185
APÊNDICE 03: GRÁFICOS DE DISPERSÃO E LINEARIDADE .....		186
APÊNDICE 04: ESTATÍSTICAS DE 50 REGRESSÕES LOCAIS DE GWR.....		187
APÊNDICE 05: ALGUNS CÓDIGOS UTILIZADOS NESSE ESTUDO .....		188
APÊNDICE 06: MODELAGEM CLÁSSICA (SEM AS VARIÁVEIS DE BAIRRO)....		190
APÊNDICE 07: SEMIVARIOGRAMAS TEÓRICOS DAS MODELAGENS .....		191

## 1 INTRODUÇÃO

Nos dias atuais, a escassez de recursos fiscais, somada às crescentes demandas sociais decorrentes do desenvolvimento urbano, tem requerido das administrações públicas grandes esforços para que as receitas e despesas dos municípios brasileiros se mantenham em equilíbrio orçamentário-financeiro.

O Imposto Predial e Territorial Urbano, conhecido pela sigla IPTU, é um tributo de competência municipal e uma importante fonte de arrecadação de recursos financeiros destinados à gestão urbana dos municípios brasileiros.

Conforme a Constituição Federal (1988), o IPTU é um imposto municipal calculado a partir do valor venal da propriedade urbana, que pode ter alíquotas diferenciadas, de acordo com a localização e utilização do imóvel. O Imposto Predial e Territorial Urbano (IPTU) tem, portanto, como principal referência de cálculo, o valor venal do imóvel, que deve corresponder ao valor de venda à vista, obtido por meio do somatório do valor do terreno com o valor da respectiva edificação.

Como a base de cálculo do IPTU é o valor venal dos imóveis, a administração municipal deve sempre determiná-lo da forma mais realista possível, fazendo com que esse valor esteja próximo do valor de mercado. Contudo, devido à complexidade para a sua determinação, bem como pelo elevado número de imóveis a serem avaliados, isso se torna uma tarefa árdua (CARVALHO JR., 2011).

Segundo Liporoni (2013), a avaliação em massa tem como objetivo a determinação, em larga escala, dos valores de imóveis, de forma a manter justa proporcionalidade destes valores entre si, em face da localização e das características específicas de terrenos, por meio de metodologia e critérios de avaliações.

No Brasil, a Associação Brasileira de Normas Técnicas (ABNT) edita a Norma Brasileira de Avaliação de Bens (NBR 14-653) composta, atualmente, por sete partes. Na parte dois (NBR 14.653-2), a norma estabelece as diretrizes para a avaliação de imóveis em áreas urbanas e recomenda, sempre que possível, o uso do método comparativo direto de dados de mercado para a avaliação de imóveis.

Conforme Cesare (2012), há pleno consenso na área de avaliações de imóveis da superioridade do método comparativo direto de dados de mercado. Entretanto, ainda é incomum a aplicação do método comparativo direto de dados de mercado para a avaliação em massa de casas, apartamentos, salas, lojas, etc nos municípios brasileiros. Em geral, seu uso é limitado à avaliação de terrenos.

Entre as técnicas comumente empregadas para a avaliação em massa está a regressão linear múltipla. Conforme Uberti (2016), tem-se verificado a necessidade de se incorporar variáveis de localização nos modelos de regressão, com o objetivo de se modelar os efeitos espaciais, atingindo assim, maior precisão e justiça fiscal.

Contudo, segundo Dantas (2003), o tratamento incompleto ou inadequado dos efeitos espaciais pode gerar sérios problemas de estimação, pois, na presença de autocorrelação espacial nos resíduos, os parâmetros estimados são ineficientes e, nesse caso, os testes de significância e os intervalos de confiança inferidos não são mais válidos e as decisões tomadas com base neles são enganosas.

Dessa forma, para minimizar os prejuízos aos modelos de regressão decorrentes dos efeitos espaciais, há técnicas de geoestatística e de regressão espacial.

O uso da econometria espacial pode ser utilizado para modelagem dos efeitos espaciais sobre o valor de mercado dos imóveis na elaboração de planta de valores genéricos (PVG) em áreas urbanas e, dentre os métodos mais utilizados pelos profissionais de engenharia de avaliações, estão os modelos de regressão do erro e da defasagem espacial, ambos, modelos globais de regressão espacial.

Contudo, outras técnicas podem ser aplicadas visando a melhora do poder de predição do modelo quando variáveis de localização são a ele incorporadas. Exemplo disso, tem-se a regressão geograficamente ponderada (RGP), em inglês *Geographically Weighted Regression* (GWR), desenvolvida por Brunsdon, Fotheringham e Charlton (1996), e utilizada para modelar processos heterogêneos (não estacionários) espacialmente, ou seja, processos que variam de região para região, fazendo com que a modelagem da tendência espacial ocorra de forma contínua, com parâmetros variando no espaço. A ideia por trás da regressão geograficamente ponderada é ajustar um modelo de regressão para cada ponto observado, ponderando todas as demais observações como função da distância a esse ponto.

Desta forma, ao se utilizar a regressão geograficamente ponderada (GWR), tem-se como vantagem a possibilidade de variação dos parâmetros do modelo conforme sua localização geográfica, ao passo que um modelo de regressão espacial global pode não representar adequadamente as variações locais.

Sendo assim, tendo como base a constatação de que a avaliação em massa de imóveis, em geral, tem seu uso limitado à avaliação de terrenos, bem como a necessidade de se realizar o tratamento adequado dos efeitos espaciais presente nos dados, esse estudo investiga a possibilidade de geração de uma planta de

valores genéricos para imóveis construídos, incorporando elementos de localização, por meio de modelos de regressão geograficamente ponderada.

Pretende-se, então, nesta pesquisa, investigar como se dá a avaliação em massa de imóveis tipo apartamento utilizando regressão geograficamente ponderada (GWR) para a geração de plantas de valores genéricos de imóveis construídos.

## 1.1 OBJETIVOS

Na busca por respostas às críticas e aos questionamentos, o presente trabalho de pesquisa verifica a possibilidade de utilização de modelos de regressão geograficamente ponderada na avaliação em massa de imóveis, cujos objetivos são:

### 1.1.1 Objetivo Geral

Como objetivo geral para essa pesquisa propõe-se investigar o uso de modelos de regressão geograficamente ponderada (GWR) como ferramenta para avaliação em massa de imóveis urbanos tipo apartamento, buscando propor um método para a elaboração de planta de valores genéricos de imóveis construídos.

### 1.1.2 Objetivos Específicos

- A) Realizar o tratamento dos efeitos espaciais presentes nos dados da amostra por meio de regressão geograficamente ponderada (GWR) e comparar os resultados com os obtidos ao se utilizar a regressão do erro/defasagem espacial (CAR/SAR) e pela inclusão de variáveis de localização na regressão linear clássica (OLS);
- B) Comparar os coeficientes e as variáveis significativas dos modelos de regressão geograficamente ponderada gerados para cada uma das funções de ponderação propostas na literatura (Gaussiana e Biquadrática), e diferentes larguras de banda, verificando se existem diferenças significativas entre esses modelos;
- C) Avaliar o desempenho da modelagem por regressão geograficamente ponderada utilizando as métricas para avaliação em massa estabelecidas pela norma IAAO;
- D) Gerar superfícies de valores para os bairros analisados, utilizando krigagem ordinária, a partir da interpolação dos valores preditos pelos modelos de regressão clássica, espacial do erro/defasagem e geograficamente ponderada.

## 1.2 DELIMITAÇÃO DA PESQUISA

Conforme Rudio (1985, p. 72), a delimitação de uma pesquisa deve ater-se à definição do campo de observação, e essa comporta, além do local (recorte espacial) e circunstância (recorte temporal), a população que será estudada.

Nesse sentido, o presente trabalho de pesquisa busca investigar o uso de modelos de regressão geograficamente ponderada (GWR) para avaliação em massa de imóveis urbanos, utilizando para esse fim 225 dados de mercado de imóveis de tipologia apartamento, coletados nos meses de março e abril de 2020, nos bairros Centro, Agronômica e Trindade, no município de Florianópolis, em Santa Catarina.

## 1.3 JUSTIFICATIVA

Concordando com Moura *et al.* (2004), no Brasil, praticamente todos os municípios de médio e grande porte possuem seus sistemas cadastrais de imóveis, mas muitos destes sistemas são, ao seu modo, incompletos e ineficientes.

Essa afirmação, mesmo nos dias atuais, é uma realidade presente no Brasil. Dessa forma, os sistemas cadastrais e as plantas de valores genéricos dos imóveis dos municípios tornam-se barreiras para a correta tributação do IPTU.

Em geral, a diferença entre o valor de mercado dos imóveis e os valores que constam na planta de valores genéricos dos municípios apresenta discrepâncias significativas. Essa discrepância se dá, especificamente, nas plantas de valores genéricos dos imóveis, que, na grande maioria das vezes, se apresentam desatualizadas, impedindo a implantação de uma política tributária com justiça fiscal.

Não fosse o bastante, Carvalho Jr. (2011) alerta que muitos municípios se utilizam do mesmo valor de imóvel para determinada zona da cidade ou até para todo o perímetro urbano, gerando distorções significativas no nível das avaliações.

Segundo Fontoura Jr. *et al.* (2018), as técnicas de avaliação usadas ainda são baseadas em grande parte na estatística descritiva (utilização de fatores), sendo que estas técnicas não têm apresentado resultados satisfatórios para determinar o valor do imóvel, pois tem como base de cálculo coeficientes e fórmulas prontas que não refletem o comportamento dos valores do mercado imobiliário.

Além disso, os modelos tradicionalmente utilizados na avaliação em massa, baseados na inferência estatística por mínimos quadrados, geralmente, apresentam

dificuldades perante a complexidade de fatores de localização, por serem extremamente amplos, desde polos valorizantes, como proximidade ao centro principal, escolas, hospitais etc, até polos que causam situações indesejáveis, como a presença de uma favela, lixão etc. (LIPORONI, 2013).

Conforme o mesmo autor, os métodos de avaliação em massa devem considerar essas características espaciais de forma adequada e a estatística espacial passa a se constituir uma ferramenta importante nesse sentido.

Contudo, Trivelloni e Hochheim (2006) alertam para o fato de que um problema comum na avaliação em massa de imóveis é a forma como é feito o tratamento da variável localização. Eles dizem que as técnicas tradicionais de inferência estatística podem levar a resultados ineficientes na existência de auto-correlação espacial nos resíduos dos modelos. Diante dessa situação, o pesquisador precisará utilizar-se de técnicas de estatística espacial para o tratamento dos dados.

Segundo Dantas (2003), o primeiro passo para a utilização de modelos espaciais é o diagnóstico da presença de efeitos espaciais. A dependência espacial tem sido diagnosticada de duas formas distintas: pela análise gráfica do variograma ou utilizando testes estatísticos específicos, que utilizam matrizes de vizinhança.

Do ponto de vista econométrico, ao se incluir efeitos espaciais no modelo de regressão, está se considerando, por exemplo, que a evolução de uma determinada região pode ser influenciada pela evolução nas regiões vizinhas. Liporoni (2013) ensina que a existência desses efeitos pode ser determinada por um conjunto de técnicas que têm sido desenvolvidas pela econometria espacial, na qual Anselin, com diversos trabalhos, tem dado grande contribuição. Segundo Anselin (1999), a econometria espacial é a parte da econometria que lida com o tratamento da estrutura espacial e as interações de dados geograficamente distribuídos.

Em diversas pesquisas de avaliação em massa de imóveis urbanos, têm sido incorporadas às metodologias avaliatórias, técnicas que permitem identificar a influência que a dependência espacial exerce sobre a formação dos preços, bem como a influência que um imóvel exerce sobre seus vizinhos (HORNBURG, 2009).

Em relação a essas técnicas que incorporam efeitos espaciais na geração de modelos de avaliação, está a regressão espacial, citada anteriormente, e a regressão geograficamente ponderada (GWR), que possibilitam que se ajuste um modelo de regressão para cada ponto do conjunto de dados, ponderando todas as demais observações, como função da distância ao respectivo ponto.

Nesse sentido, este trabalho de pesquisa se propõe a aplicar a técnica de regressão geograficamente ponderada (GWR) na avaliação em massa de imóveis tipo apartamento em três bairros da cidade de Florianópolis e investigar a obtenção de modelos mais eficientes e precisos, que considerem os efeitos espaciais.

Com o desenvolvimento desse trabalho, pretende-se gerar conhecimentos que contribuirão para a solução de problemas relacionados à avaliação em massa de imóveis urbanos para geração de plantas de valores genéricos para prefeituras.

Então, a motivação e a justificativa para o desenvolvimento deste trabalho de pesquisa decorrem da necessidade iminente de se ter um método mais preciso e, ao mesmo tempo justo, de avaliação em massa de imóveis para fins fiscais.

#### 1.4 ESTADO DE ARTE DA GWR NA AVALIAÇÃO DE IMÓVEIS

É consenso que o elemento localização é um dos principais elementos de valorização ou desvalorização de um imóvel. Conforme Trivelloni (2005), a dependência espacial existente no valor dos imóveis, produzida principalmente pelos fatores de localização, tem mostrado que as técnicas tradicionais de inferência estatística podem levar a resultados ineficientes na presença de autocorrelação espacial dos resíduos do modelo. Ele diz, ainda, que duas grandes escolas metodológicas de estatística espacial têm sido propostas: a que propõe usar métodos de econometria espacial e a que propõe usar métodos geoestatísticos.

Dubin (2003) comparou os métodos econométricos espaciais e os modelos geoestatísticos por meio de uma aplicação prática que analisou os resultados em função do grau de ajustamento entre os valores observados e calculados por ambos os métodos e a análise dos erros obtidos sobre uma mesma amostra de dados, aplicando as duas metodologias. O autor conclui que os modelos geoestatísticos apresentam superioridade na estimação por inferência estatística sobre os modelos de regressão espacial, sendo que ambos os métodos, por sua vez, são reconhecidamente superiores aos resultados obtidos por regressão linear múltipla.

As metodologias de econometria espacial e geoestatística foram utilizadas de forma combinada em uma amostra retirada de bairros do município de São José (SC), por Trivelloni (2005). Os modelos de regressão espacial e os métodos geoestatísticos foram utilizados para estimar o valor da localização dos imóveis. Entre as conclusões, destaca-se o fato de a utilização combinada das duas técnicas

terem apresentado resultados superiores àqueles obtidos de forma individualizada, possibilitando a análise de uma amostra heterogênea de dados e a determinação do valor da localização. O autor utilizou os resultados obtidos para gerar uma planta de valores genéricos (PVG) para a área em estudo.

Contudo, ainda são insipientes no Brasil as pesquisas que utilizam a regressão geograficamente ponderada na avaliação de imóveis para modelar a não estacionaridade (heterogeneidade) espacial, condição na qual a adoção de um modelo de regressão espacial global não conseguiria explicar adequadamente as relações entre algumas variáveis definidas em determinada região geográfica.

Legg *et al.* (2009) aplicaram os conceitos de regressão geograficamente ponderada a um problema imobiliário em Michigan. Para geração do modelo, utilizaram dados de 93 casas e, entre os dados, estavam as coordenadas UTM. Inicialmente, a equipe gerou um modelo de regressão espacial e, na sequência, um modelo de regressão geograficamente ponderada foi gerado com o mesmo conjunto de dados. Eles constataram que, embora o modelo de regressão linear tenha sido significativo e tenha um forte valor de  $R^2 = 0,782$  ( $p = 0,000$ ), o modelo GWR melhorou essas estatísticas e aumentou a precisão do modelo para um valor de  $R^2 = 0,865$  ( $p = 0,000$ ). A equipe concluiu que a ferramenta GWR é de simples utilização e oferece uma base sofisticada para quantificar e dissecar padrões espaciais, assim como perceberam que a aplicação da GWR oferece uma melhoria notável em relação à regressão linear, aumentando, assim, a precisão da modelagem.

Uma modelagem de aluguéis residenciais considerando os efeitos espaciais foi realizada por Lochl e Axhausen (2010). Um conjunto de dados abrangente é usado para modelar as solicitações de aluguéis residenciais, comparando técnicas de regressão clássica, espacial e geograficamente ponderada. Os autores concluem que, embora todas as medidas de qualidade indicassem um desempenho ligeiramente melhor para o modelo obtido por GWR em comparação com os demais, seus resíduos ainda se apresentavam autocorrelacionados.

Jahanshiri *et al.* (2011) fizeram uma revisão sobre modelos de avaliação em massa de imóveis, entre eles a regressão linear clássica e vários outros métodos avançados, como SAR (regressão espacial) e *Kriging*, bons para modelagem da dependência espacial ou GWR (regressão geograficamente ponderada) e MWR (regressão de janelas móveis), bons para modelagem da heterogeneidade espacial.

O impacto de parques eólicos nos valores de propriedades, utilizando um modelo de preços hedônicos geograficamente ponderado, foi avaliado por Sunak e Madlener (2012). Os autores avaliaram o impacto provocado pela implantação de projetos de energia eólica na região da Renânia do Norte (Vestfália, Alemanha) sobre os preços das terras, usando a regressão geograficamente ponderada (GWR).

O uso de regressão geograficamente ponderada para analisar o mercado imobiliário foi abordado por Celmmer (2012). O autor utilizou as técnicas de GWR para analisar os preços das transações no mercado imobiliário e avaliar o efeito de atributos de propriedades no valor dos imóveis. Os dados de mercado foram analisados para avaliar a influência da disponibilidade de infraestrutura nos preços de imóveis em Olsztyn, na Polônia. A avaliação do efeito dos serviços de utilidade pública nos preços dos imóveis indica que a regressão geograficamente ponderada é uma ferramenta útil para as análises reais do mercado. Segundo o autor, a técnica de GWR ajudou a avaliar o efeito dos atributos sobre os preços dos imóveis.

Lockwood e Rossini (2012) calcularam a eficácia do modelo de regressão geograficamente ponderada no processo de avaliação em massa. O trabalho compara o uso de um modelo hedônico de regressão geograficamente ponderada (GWR) com os modelos hedônicos mais tradicionais. Enquanto todos os modelos gerados apresentaram estatísticas dentro de níveis aceitáveis de precisão, a capacidade de prever valor de terrenos demonstrou resultados mais interessantes para o modelo GWR. Segundo os autores, o uso de um *kernel* fixo em diferentes larguras de banda pode ajudar a entender ainda mais o uso do conceito GWR.

Uma avaliação da precisão do modelo espacial na avaliação imobiliária em massa foi realizada por Bidanset e Lombard (2014). Os autores compararam o uso da regressão geograficamente ponderada (GWR) e do modelo de defasagem espacial (SAR). O trabalho é realizado usando vendas residenciais de 2010 a 2012 em Norfolk, Virgínia, e compara o desempenho da GWR e SAR, extrapolando, assim, o desempenho de cada modelo para agregar e subagregar níveis. Constatou-se que a GWR aumenta significativamente o desempenho de modelos de avaliação. Resultados indicam que a GWR atinge coeficiente de dispersão menor que o SAR.

Dimopoulos e Moulas (2016) propuseram um sistema de avaliação em massa na Grécia com uso de um sistema assistido por computador (CAMA), com o qual avaliaram as técnicas de regressão geograficamente ponderada (GWR) e a análise de regressão linear múltipla (RLM) no município de Thessaloniki. Os autores

utilizaram 90% dos dados para geração dos modelos e 10% dos dados para verificação dos resultados. Os dois modelos passaram em todas as verificações estatísticas e têm excelente precisão. No entanto, verificaram que o método GWR possui várias vantagens sobre o método OLS. O método GWR tem uma eficiência muito mais alta na determinação de  $R^2$ . O valor  $AICc$  do modelo GWR é menor que o do OLS. Por todas as razões acima, os autores concluem que o método GWR pode levar a modelos de previsão superiores comparados com o método OLS tradicional.

O método de regressão geograficamente ponderada (GWR) foi utilizado por Dziauddin e Idris (2017) para estimar os efeitos dos atributos de localização nos valores das propriedades residenciais. O estudo estimou o efeito dos atributos de localização nos valores das propriedades residenciais em Kuala Lumpur, na Malásia. Os resultados do estudo revelaram que os valores das propriedades residenciais são determinados, principalmente, pelos atributos físicos (estruturais) da propriedade, mas a proximidade com os atributos locais também contribui marginalmente. O uso da GWR foi considerado uma abordagem melhor do que outros métodos para examinar o efeito dos atributos locacionais nos valores de propriedades residenciais.

Além disso, Flores e García (2018) usaram regressão geograficamente ponderada para analisar a influência da componente espacial na avaliação de propriedades residenciais. Selecionou-se como área de estudo o município de Caguas, na ilha de Porto Rico. A pesquisa procurou mostrar como a introdução da variabilidade espacial melhora a previsibilidade na modelagem de valores imobiliários. Os resultados mostram que a regressão geograficamente ponderada (GWR) apresenta um desempenho superior quando comparada à modelagem de regressão clássica por mínimos quadrados ordinários.

Uberti (2016), mais recentemente, utilizou, além da modelagem clássica e espacial, a regressão geograficamente ponderada para avaliação em massa de imóveis rurais. O trabalho consistiu em investigar a metodologia estatística mais adequada para a elaboração de PVG para áreas rurais utilizando dados e mapas do PROJIR (Projeto de Irrigação e Drenagem na Região Norte Fluminense). Para o modelo de regressão geograficamente ponderada (GWR), foram testados os *kernels* fixo e adaptativo, sendo que o *kernel* biquadrado de largura fixa foi o que apresentou o menor valor para o critério de Akaike ( $AIC$ ). Na sequência, os três modelos foram avaliados na qualidade do ajuste, utilizando-se do critério de Akaike, a função de verossimilhança e o coeficiente de determinação, assim como no desempenho do

ajuste, utilizando-se de 25 dados de amostra de verificação. Constatou-se, então, que o modelo de regressão geograficamente ponderada foi superior ao modelo de regressão clássica e de regressão espacial em todos os critérios, deixando a modelagem mais ajustada em relação aos outros dois modelos.

## 1.5 ESTRUTURA DA PESQUISA

O presente estudo está organizado em sete (7) capítulos, da seguinte forma:

No Capítulo 1, apresenta uma breve introdução sobre o trabalho, sua proposta, contextualiza e justifica o tema abordado evidenciando sua importância, define os objetivos e apresenta a estrutura e a metodologia utilizada no trabalho.

No Capítulo 2, é realizada uma revisão bibliográfica envolvendo engenharia de avaliação de imóveis, abordando temas como avaliação em massa de imóveis para fins tributários, planta de valores genéricos e cadastro técnico multifinalitário.

No Capítulo 3, é efetivada uma revisão da literatura relacionada à regressão clássica e espacial, em que se aborda a obtenção de modelos pelo método dos mínimos quadrados ordinários e pelo método do erro e defasagem espaciais.

No Capítulo 4, é realizada uma revisão bibliográfica envolvendo a regressão geograficamente ponderada, abordando a concepção do modelo de GWR, as funções de ponderação espacial, o processo de calibração do modelo, os testes aplicados à regressão geoponderada e algumas medidas de qualidade do ajuste.

No Capítulo 5, é fundamentada a metodologia e apresentados os métodos, a classificação da pesquisa e o roteiro metodológico adotado neste trabalho. Neste capítulo é exposto, ainda, a descrição dos dados que compõem a amostra.

No Capítulo 6, são apresentados os resultados da pesquisa referentes à análise exploratória dos dados e aplicação dos diferentes modelos de regressão (regressão clássica, espacial e geograficamente ponderada); são realizados os cálculos referentes à qualidade e desempenho dos ajustes e, então, os resultados obtidos são avaliados quanto à possibilidade de geração de plantas de valores.

No Capítulo 7, é realizado o fechamento deste trabalho, apresentando as considerações finais, as limitações e as recomendações para trabalhos futuros.

## 2 ENGENHARIA DE AVALIAÇÕES DE IMÓVEIS

A engenharia de avaliações é o ramo da engenharia responsável pela avaliação objetiva de imóveis. Segundo Dantas (2012, p. 25), a engenharia de avaliações é uma especialidade da engenharia que reúne um conjunto amplo de conhecimentos na área de engenharia e arquitetura com o objetivo de determinar tecnicamente o valor de um bem, de seus direitos, frutos e custos de reprodução.

Nesse sentido, a avaliação de imóveis se apresenta como uma grande atividade da engenharia, com impacto direto na viabilização de milhares de negócios mundo afora. Os primeiros trabalhos de engenharia de avaliações que se têm conhecimento remontam ao início do século XX, com destaque ao engenheiro Berrini (DANTAS, 2012), um dos precursores dessa área no Brasil.

Considerando as atuais pesquisas sobre engenharia de avaliações, pode-se verificar que essa atividade se apresenta em um avançado estágio de evolução e um importante elemento dessa evolução é a NBR 14.653 em suas partes de 1 a 7, possibilitando que procedimentos sejam realizados de forma padronizada e introduz os conceitos e as exigências técnicas para a avaliação de bens em geral, realizada com fundamentação e precisão mínimas desejadas.

Contudo, mesmo com a adoção de ferramentas matemáticas sofisticadas e poderosos *softwares* voltados para a área de engenharia de avaliações, sabe-se, ainda, que há muita imprecisão inerente ao próprio ato de avaliar.

Existem vários métodos disponíveis para avaliação de imóveis, previstos na NBR 14.653:1 - Norma de Avaliação de Bens, que podem ser empregados para estimar o valor de um bem, de seus frutos e direitos, o seu custo ou ainda determinar indicadores de viabilidade, sendo eles: o método comparativo direto de dados de mercado, o método involutivo, o método evolutivo, o método da capitalização da renda, o método da quantificação e comparativo direto de custo.

Em geral, o valor que se quer encontrar em uma avaliação imobiliária é o valor de mercado, ou valor venal, referente à venda, definido na NBR 14653:1 – Procedimentos Gerais como o “[...] valor de mercado: quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente”. A mesma norma, em seu item 8.2, define os seguintes métodos para identificação do valor de imóveis urbanos:

- A) Método comparativo direto de dados de mercado: identifica o valor do bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra;
- B) Método involutivo: identifica o valor de mercado do bem, alicerçado no seu aproveitamento eficiente, baseado em modelo de estudo de viabilidade técnico econômica, mediante hipotético empreendimento compatível com as características do bem e com as condições do mercado no qual está inserido, considerando-se cenários viáveis de execução e comercialização;
- C) Método evolutivo: identifica o valor do bem pelo somatório das parcelas componentes dele (valor da edificação e do terreno);
- D) Método da capitalização da renda: identifica o valor do bem com base na capitalização da renda líquida prevista, considerando-se cenários viáveis;
- E) Método comparativo direto de custo: identifica o custo do bem por meio de tratamento técnico dos atributos dos elementos comparáveis da amostra;
- F) Método da quantificação de custo: Identifica o custo do bem ou de suas partes por meio de orçamentos sintéticos ou analíticos, a partir das quantidades de serviços e respectivos custos diretos e indiretos.

Verifica-se que os únicos métodos diretos são o comparativo de dados de mercado e o comparativo de custo, sendo os demais, indiretos. Um método é considerado direto quando, pela sua aplicação, obtém-se o resultado da avaliação, sem depender de qualquer outro; entretanto, a avaliação pelos métodos indiretos exige, de alguma forma, resultados dos métodos diretos (DANTAS, 2012, p.15).

A NBR 14.653-2 (2011), em seus anexos de A até E, traz, ainda, recomendações para a utilização de modelos de regressão linear, utilização de tratamento por fatores, tratamento de dados por regressão espacial, utilização de análise envoltória de dados e tratamento de dados por redes neurais artificiais.

Frente a isso, a NBR 14.653:1 recomenda que se utilize, sempre que possível, o método comparativo de dados de mercado, entretanto, em avaliações em massa de imóveis, de modo geral, busca-se fazer uma conjugação de métodos.

Segundo Afonso *et al.* (2014), na grande maioria dos municípios brasileiros, o cálculo do valor venal é resultado da adição do valor do terreno com o da construção. Esse resultado é obtido por meio do método evolutivo, que utiliza o método comparativo direto de dados de mercado para determinar o valor do terreno e o método da quantificação do custo para determinar o valor da edificação.

## 2.1 AVALIAÇÃO EM MASSA DE IMÓVEIS

A origem das avaliações em massa, segundo Liporoni (2013), confunde-se, historicamente, com a cobrança de impostos sobre a propriedade imobiliária, sendo este tributo de origem milenar. Tratando-se de uma das fontes de tributação das mais tradicionais e tendo o valor dos imóveis como base de lançamento, é certo que a evolução das metodologias de avaliação em massa foi direcionada no sentido da apuração dos valores territoriais e prediais, denominados valores venais.

De acordo com McCluskey *et al.* (1997), as avaliações de imóveis em massa e individuais são diferentes somente na escala, porque o que se busca nos dois casos é uma avaliação precisa do valor de um ou mais imóveis. Segundo esses autores, a avaliação em massa surgiu pela necessidade de uniformidade e coerência nas avaliações quando existem muitos imóveis para avaliar.

### 2.1.1 Conceitos e Métodos de Avaliação em Massa

A avaliação em massa de imóveis consiste no desenvolvimento de um ou mais modelos genéricos, originados, em geral, com base em uma amostra de dados, os quais são aplicados tomando em conta dados cadastrais. Esses modelos devem ser construídos com base em métodos normatizados e análises estatísticas, ou outras técnicas capazes de estimar com acurácia o valor dos bens (CESARE, 2012).

A avaliação em massa consiste em determinar os valores de todos os imóveis de uma região ou município. Uma importante aplicação das avaliações de imóveis é na tributação imobiliária. A correta estimação dos valores é fundamental para atingir a equidade (mesma relação para todos os imóveis entre o valor da avaliação e o valor de mercado) e para que ocorra justiça fiscal (UBERTI, 2016).

Dessa forma, a avaliação em massa de imóveis urbanos tem como objetivo a determinação sistemática, em larga escala, dos valores dos imóveis, de forma a manter uma justa proporcionalidade destes valores entre si, em face da localização de forma genérica e das características específicas de terrenos e benfeitorias, por meio de metodologia e critérios de avaliações (LIPORONI, 2013).

Frente a isso, Möller (1995) ressalta que a avaliação em massa de imóveis deve ser vinculada ao estudo do Código Tributário Municipal - CTM no que se refere

aos impostos sobre a propriedade (IPTU e ITBI, impostos da esfera municipal). Dessa forma, o resultado da avaliação se adequará às exigências legais.

Liporoni (2013) nos ensina que, após a coleta de uma amostra representativa da região, deve-se identificar as variáveis influenciantes e estabelecer as situações paradigmas da região em estudo para, desta forma, fazer o tratamento da amostra coletada e, obviamente, fazer a análise das características observadas.

Tal tratamento poderá ser realizado pela metodologia determinística (homogeneização dos valores por estatística descritiva ou por fatores de ponderação), pela metodologia probabilística (regressão linear múltipla por inferência estatística) ou, ainda, por outras técnicas, sendo aplicada aquela que melhor se adequar ao universo amostral obtido e à massa de imóveis sob avaliação.

Dessa forma, Liporoni (2013) diferencia o valor territorial do valor predial:

- **Valor Territorial:** a determinação do valor territorial de cada imóvel resultará do produto entre sua área superficial de terreno e o valor unitário genérico respectivo à face de quadra onde está localizado tal imóvel, corrigido conforme suas características específicas pelos critérios estabelecidos pelos modelos determinísticos ou probabilísticos empregados; e

- **Valor Predial:** o valor de cada edificação é definido pelo produto entre o valor unitário genérico para a tipologia e o padrão construtivo observados, conforme toda a metodologia de classificação estabelecida, multiplicado pela área edificada dos imóveis, definida por meio do banco de dados do cadastro técnico municipal ou de um cadastro específico; e, ainda, a sua respectiva depreciação, definida por critérios que considerem sua idade real, vida útil e estado de conservação ou sua idade aparente, abrangendo todos os imóveis da região.

Liporoni (2013) ressalta ainda que em toda avaliação em massa existem casos excepcionais, ou seja, imóveis cujas características destoam ou não se enquadram à massa representativa dos imóveis do município ou região avaliada, seja por extensão de área territorial ou edificada, pelo porte das construções incorporadas ou pela situação em que se localizam, além de outros condicionantes.

Dessa forma, algumas características das avaliações em massa podem ser definidas, entre elas: A) a proporcionalidade entre os valores aplicados; B) a uniformidade nos critérios de avaliação; C) a alta produtividade nos processos avaliatórios; e D) o tratamento específico para imóveis em condição excepcional.

## 2.1.2 Análise de Desempenho da Avaliação em Massa

Para se analisar o desempenho de uma avaliação em massa de imóveis, sugere-se verificar se esta cumpre as condições da *International Association of Assessing Officers* (IAAO), avaliando-se, dessa forma, a qualidade do ajuste.

O desempenho das avaliações em massa consiste basicamente em se comparar os valores obtidos na avaliação com os valores praticados no mercado, ou seja, o valor de mercado dos imóveis. Por ser um conceito bastante subjetivo, normalmente se utiliza os preços de venda para fins de análise de desempenho.

Conforme Trivelloni (2005), quanto mais próximos estejam os valores calculados pelo modelo dos valores observados no mercado, melhor será a qualidade da avaliação. Geralmente, as medidas de desempenho referem-se pelo menos a dois tipos de qualidade: medidas de desempenho global da avaliação e medidas relacionadas com a equidade horizontal e uniformidade da avaliação.

O parâmetro mais usado como medida de desempenho global de uma avaliação, segundo Cesare (2012), é a mediana das razões de avaliação (Med R), que calcula a mediana dos quocientes entre o valor avaliado e o valor de venda de determinado imóvel. Em relação aos parâmetros de dispersão da avaliação, a medida mais comumente usada para avaliar a uniformidade é o chamado coeficiente de dispersão da mediana (COD). Nesse caso, ambas as medidas são aplicáveis mesmo quando a variável em análise não seja normalmente distribuída.

Para medir a equidade vertical da avaliação, é utilizado o índice conhecido como diferencial relativo ao preço ou PRD, que serve para detectar diferenças sistemáticas na forma que são avaliados os imóveis de alto e de baixo valor, verificando se existe regressividade ou progressividade na avaliação (IAAO, 2013).

### - Mediana das Razões de Avaliação

Conforme a *Association of Assessing Officers* (IAAO, 2013), o parâmetro mais usado como medida de desempenho global de avaliações é a mediana das razões de avaliação. Dada uma observação do preço de venda ( $P_v$ ) e valor calculado pelo modelo ( $P_c$ ), chama-se razão de avaliação ( $R$ ) ou *ratio* de avaliação o quociente:

$$R = \frac{P_c}{P_v} \quad (2.1)$$

A mediana indica, segundo Cesare (2012), o nível pelo qual um grupo de imóveis foi avaliado em relação aos valores praticados no mercado de imóveis, demonstrando quão próximos os imóveis foram avaliados do nível legal desejado.

A mediana das razões de avaliação é menos afetada por relações extremas do que as outras medidas de tendência central e, devido a esta propriedade, é a medida adotada pela IAAO (2013) para avaliar o nível geral da avaliação.

Com relação aos valores recomendados para a mediana das razões de avaliação, é considerado aceitável quando seu valor está entre 90% e 110%.

#### **- Coeficiente de Dispersão da Mediana**

O coeficiente de dispersão da mediana (COD) é o desvio médio expresso em termos percentuais do nível em que cada propriedade foi avaliada em relação à mediana do valor avaliado dividido pelo valor de mercado. O coeficiente mede a variabilidade (grau de uniformidade) das respectivas avaliações (CESARE, 2012).

Segundo a IAAO (2013), o COD é obtido fazendo: A) diminui-se de cada razão de avaliação a mediana de todas as razões; B) é tomado o valor absoluto das diferenças anteriores; C) calcula-se a média dos valores absolutos; D) divide-se a média anterior pela mediana das razões de avaliação; E) multiplica-se por 100.

Com relação aos valores recomendados, as normas internacionais estabelecidas pelo IAAO (2013) estabelecem limites máximos para o coeficiente COD de 10% para imóveis residenciais novos ou em áreas homogêneas, 15% para imóveis antigos ou em áreas heterogêneas e 20% para as zonas rurais e terrenos.

Considerando que a principal finalidade da avaliação em massa é a de se obter uma base de cálculo atualizada para cobrança dos tributos imobiliários, mais especificamente, o IPTU e o ITBI, de maneira a garantir a equidade fiscal e a prática da justiça tributária, é imprescindível obter equações abrangentes e esclarecedoras, de modo que o modelo adotado permita representar o universo em estudo, simplificando a uma estrutura de variáveis intervenientes consideradas importantes e que possibilitem ser aplicadas para todos os imóveis da cidade (LIPORONI, 2013).

#### **- Diferencial Relativo ao Preço**

Conforme visto, o diferencial relativo ao preço, do inglês *Price Related Differential* (PRD), trata-se de um indicador utilizado para medir a equidade vertical da avaliação e serve para se verificar diferenças sistemáticas na forma que são

avaliados os imóveis de valores mais altos e os imóveis de valores mais baixos, possibilitando compreender se há regressividade ou progressividade na avaliação.

Segundo Theodoro *et al.* (2019), a regressividade ocorre quando os imóveis de baixo valor são avaliados como uma porcentagem maior do valor de mercado que os imóveis de alto valor, o contrário disso é a progressividade.

Para se obter o diferencial relativo ao preço, deve-se dividir a média das razões de avaliação pela média ponderada das razões de avaliação. A IAAO (2013) recomenda que o valor do PRD esteja entre 0,98 e 1,03, inclusive. Se o valor do PRD for menor que 0,98, está caracterizado a progressividade na avaliação, e se o valor do PRD for maior que 1,03, está caracterizado a regressividade na avaliação.

Cesare (2012) diz que as distorções de caráter avaliatório são responsáveis por gerar, além de outros problemas, a regressividade na cobrança do imposto, na medida em que propriedades de alto valor são subavaliadas, em termos relativos, em comparação com avaliações de imóveis de baixo valor.

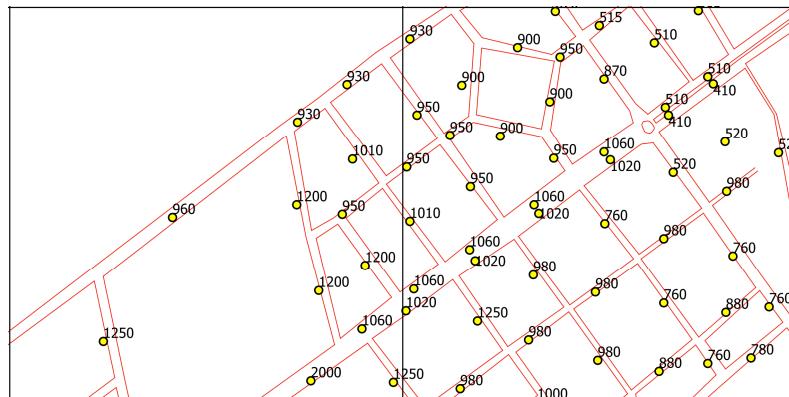
A autora afirma, ainda, que o baixo desempenho nas avaliações realizadas para fins tributários pode estar relacionado ao limitado acesso às informações de mercado, ao emprego de métodos e/ou técnicas avaliatórias não apropriadas, ao longo intervalo entre reavaliações genéricas e à omissão de atributos importantes.

## 2.2 PLANTA DE VALORES GENÉRICOS (PVG)

Baleiro (2016, p. 379), compreendendo que a tarefa de avaliar um a um os imóveis é extremamente onerosa, esclarece que as administrações municipais passaram a utilizar-se de planta de valores, documentos em que são estabelecidos certos critérios técnicos para a determinação do valor do imóvel, a partir do qual se estabelece um valor para o metro quadrado e se determina o valor do imóvel.

Conforme a NBR 14.653-2 (2011), a planta de valores genéricos (PVG) é a representação gráfica ou listagem dos valores genéricos do metro quadrado de terreno ou do imóvel em uma mesma data. Hornburg (2015) expõe que a PVG consiste em um documento gráfico que representa a distribuição espacial dos valores médios dos imóveis em cada região da cidade, normalmente apresentados por face de quadra como mostra a Figura 1. Sua principal função é permitir a definição de uma política de tributação imobiliária, que seja justa e tenha equidade.

Figura 1: Recorte de uma PVG com o valor unitário por face de quadra



Fonte: o autor

Complementando tal ideia, Liporoni (2013) afirma que a PVG deve apresentar valores médios unitários de terrenos para cada face de quadra do município. Sua elaboração, por isso mesmo, constitui-se de um trabalho bastante extenso, que, nem por isso, deve prescindir de um grande nível de detalhamento, uma vez que todas as peculiaridades de cada local devem ser consideradas.

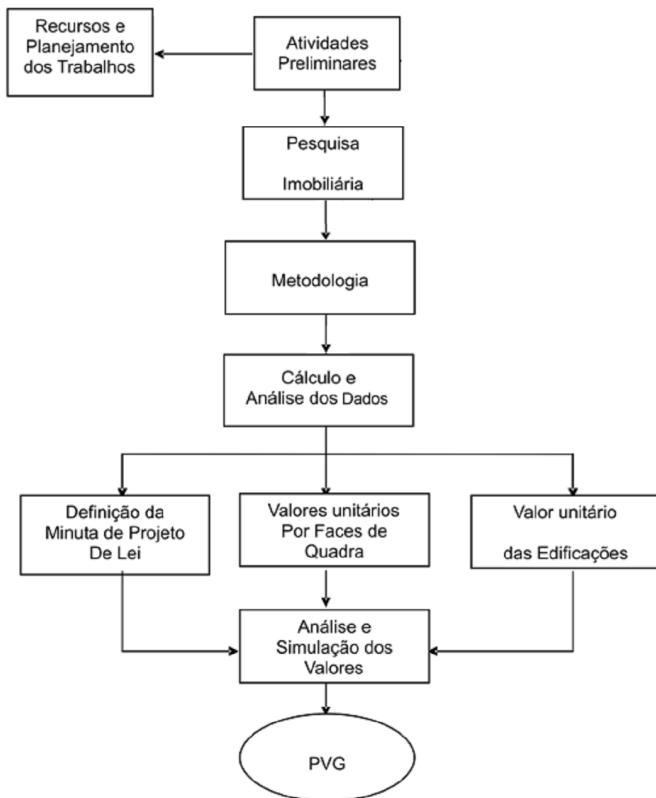
Conforme Dualde (2004), na sua denominação, a planta de valores genéricos expressa o termo genérico aplicado ao tipo de valor a que esse instrumento fiscal faz referência. A PVG é aplicada como instrumento de avaliação em massa com o objetivo de oferecer elementos para o cálculo do valor venal.

Nesse sentido, a principal função da planta de valores genéricos é permitir a definição de uma política de tributação imobiliária que seja acima de tudo justa e equânime. Assim, o processo de avaliação de imóveis para elaboração de uma PVG exige a compreensão das características básicas da população de imóveis da cidade de modo que o modelo adotado permita a avaliação de todos eles, salvo exceções, pois, ao final, deve-se obter o valor individual de cada um dos imóveis.

Para se elaborar uma PVG, pode-se adotar a homogeneização de valores, por meio de modelos determinísticos - o mais comumente utilizado – ou por meio da inferência estatística por meio de modelos probabilísticos.

Segundo Liporoni (2013), a característica principal para a elaboração da planta de valores, apesar das peculiaridades de cada município, é manter sua uniformidade no estabelecimento dos valores unitários médios de cada face de quadra. Portanto, a organização de um projeto de planta de valores é fundamental para atingir este objetivo, sugerindo-se fluxograma da Figura 2 para sua elaboração:

Figura 2: Fluxograma para elaboração de uma planta de valores



Fonte: adaptada de Liporoni (2013)

Os valores venais contidos na PVG são multiplicados pela área do terreno/edificação e em alguns municípios ainda são atribuídos coeficientes de ponderação, os quais consideram a área construída e/ou o padrão construtivo e/ou as características dos terrenos. Assim, é importante que o valor venal seja o mais próximo possível do valor de mercado dos imóveis. Entretanto, no Brasil, isso está longe de se tornar realidade, uma vez que, por motivos diversos, os valores venais são bem diferentes dos valores de mercado (CARVALHO JR., 2011).

Conforme Liporoni (2013), a planta de valores genéricos é parte integrante e básica do sistema de informações do cadastro municipal e, juntamente ao cadastro imobiliário, forma a base de cálculo do IPTU, do ITBI e da contribuição de melhoria.

### 2.2.1 Cadastro Territorial Multifinalitário (CTM)

Conforme Vasconcellos (2001), cadastro é um inventário da propriedade de imóveis em seus aspectos físicos, jurídicos e econômico-fiscais, devidamente

classificado. O autor acrescenta que a missão do cadastro é fornecer informações sobre a riqueza do território para contribuir com o desenvolvimento do país.

O cadastro territorial multifinalitário (CTM), em nível urbano ou rural, segundo Lima *et al.* (2000), é um conjunto de informações gráficas, descritivas e tabulares de uma porção da superfície terrestre, contendo as propriedades imobiliárias corretamente georreferenciadas, possibilitando o conhecimento sobre todos os aspectos.

Em função do dinamismo urbano, que acaba por ocasionar alterações significativas nos imóveis, deve existir uma estrutura que estimule a constante atualização do cadastro, sob possibilidade de, mesmo com atualizações dos valores venais dos imóveis, não se obter a adequada justiça tributária.

Ramos *et al.* (2002) ressaltam a importância do cadastro no processo de avaliação de imóveis, de modo a garantir uma maior qualidade nas ações públicas ligadas ao valor dos imóveis, isto é, tributação, desapropriações e planejamento.

Os autores propõem, ainda, a integração da base de dados do cadastro com base de dados do mercado imobiliário, para o desenvolvimento frequente de modelos de avaliação com o objetivo de manter atualizada a planta de valores.

Entre esses dados, destaca-se a identificação dos proprietários dos imóveis, as áreas e os valores venais de terrenos e edificações. É importante salientar, ainda, que o processamento desses dados, conforme a legislação que institui a planta valores genéricos e a legislação tributária municipal vigente, resulta nos valores de lançamento de impostos e taxas de base imobiliária, como o IPTU e o ITBI.

Nesse sentido, a utilização de uma planta de valores genéricos está diretamente relacionada à existência de um cadastro de imóveis de boa qualidade, que, em sua origem, era uma simples listagem de dados sobre os imóveis do município, e hoje é um sistema complexo e importante para a adequada caracterização das unidades, dos indivíduos e de seus elementos, integrando informações relacionadas à educação, saúde, ao saneamento, aos tributos e outros.

## **2.2.2 Imposto Predial e Territorial Urbano (IPTU)**

A instituição de um tributo é sempre feita mediante a lei, e sua arrecadação e fiscalização constituem atividades administrativas sempre vinculadas, ou seja, segundo nos ensina Mello (2015), devem ser cobrados nos casos previstos em lei, no prazo determinado em lei e de quem a lei determina que se exija o tributo.

A definição de tributo está prevista no art. 3º do Código Tributário Nacional: “tributo é toda prestação pecuniária compulsória, em moeda ou cujo valor nela se possa exprimir, que não constitua sanção de ato ilícito, instituída em lei”.

Regulamentado pelo Código Tributário Nacional (Lei 5.172/66), o imposto predial e territorial urbano (IPTU) é o tributo sobre a propriedade predial e territorial urbana e tem como fato gerador a propriedade, o domínio útil ou a posse de bem imóvel por natureza ou por acessão física, como definido na lei civil, localizado na zona urbana do município. Como área urbana, a lei define se tratar de área que possua dois ou mais melhoramentos dentre os definidos em lei.

O momento de ocorrência do fato gerador do imposto predial e territorial urbano se dá no dia primeiro de janeiro de cada ano. O fato gerador é a propriedade, o domínio útil ou a posse de bem imóvel por natureza ou por acessão física. Com relação à base de cálculo, esta encontra-se definida pelo art. 33º do Código Tributário Nacional, ou seja, a base de cálculo é o valor venal do imóvel.

Conforme Sá (2014), o valor venal da propriedade imobiliária é considerado como o preço mais provável pelo qual um imóvel poderia ser vendido em condições normais de negociação. Geralmente, ele é calculado pela soma dos valores venais do terreno e da edificação. O valor venal do terreno é determinado com mais facilidade quando existe uma planta de valores genéricos no município, pois, por meio dela, consegue-se estabelecer o valor unitário do terreno de cada logradouro.

Concordando com Tristão (2003), a base de cálculo do IPTU é o valor venal do imóvel, que é estimado pela administração municipal. É possível a determinação individual do valor do imóvel com a utilização de sistemas de avaliação em massa. Assim, comumente são utilizados métodos de avaliação em massa que irão gerar a planta de valores genéricos para os imóveis do município. Espera-se que a PVG possa refletir a realidade, apontando o valor unitário, em R\$/m<sup>2</sup>, dos terrenos do município para que, então, possa-se determinar o valor venal do imóvel.

Sendo assim, o valor venal de um terreno pode ser obtido fazendo o produto do valor unitário, em R\$/m<sup>2</sup>, constante na planta de valores genéricos, pela área do respectivo terreno. Já o valor venal da edificação é baseado no tipo e padrão construtivo do imóvel. Tanto o valor do terreno quanto o da edificação sofrem ajustes para considerar as suas características particulares. Portanto, a soma do valor venal do terreno com o valor venal da edificação proporcionará o valor do imóvel, que, sobre ele, irá incidir as alíquotas definidas pelo município.

### 3 MODELOS DE REGRESSÃO CLÁSSICA E ESPACIAL

É ainda um desafio a ser superado a avaliação em massa de imóveis para fins fiscais. Em geral, para que a avaliação seja confiável, as ferramentas que serão utilizadas precisam estar adaptadas à natureza do problema e os resultados obtidos precisam ser fundamentados em parâmetros de desempenhos estabelecidos na literatura científica. No tópico seguinte será apresentada uma breve revisão das duas principais metodologias utilizadas na avaliação em massa de imóveis: o modelo clássico de regressão linear e os modelos de regressão espacial.

Conforme Gujarati e Porter (2011), em termos formais, a regressão diz respeito ao estudo da dependência de uma variável, a variável dependente ou explicada, em relação a uma ou mais variáveis explanatórias ou explicativas, visando estimar e/ou prever o valor médio (da população) da primeira em termos dos valores conhecidos ou fiados (em amostragens repetidas) das segundas.

#### 3.1 MODELO DE REGRESSÃO CLÁSSICA

Os modelos de regressão são amplamente usados em diversas aplicações para descrever a relação entre a variável resposta  $y$  e uma ou mais variáveis explicativas  $x_1, x_2, \dots, x_m$ . O modelo de regressão é linear se a quantidade de interesse é função linear dos parâmetros, caso contrário, o modelo é não-linear.

O modelo clássico de regressão linear é ainda uma ferramenta bastante utilizada em diversas áreas. Utilizando as técnicas de regressão, pode-se estabelecer uma relação linear entre a variável explicada, por exemplo o valor venal de um imóvel, e as variáveis explicativas, como sua área ou seu padrão construtivo.

Trata-se de um modelo simples, com hipóteses bastante restritivas, mas que os parâmetros de desempenho são conhecidos. Segundo Serbanouiu (2019), é uma ferramenta que é também amplamente usada pelo mercado imobiliário e que pode ser flexibilizada ao permitir que as próprias variáveis independentes sejam objeto de transformações. O autor ensina, por exemplo, que se pode estipular que o valor venal do imóvel depende linearmente da área construída e do quadrado do padrão de acabamento, buscando sempre o atendimento aos pressupostos básicos do modelo de regressão. Essas transformações são possibilidades para se alcançar a independência dos resíduos gerados pelo modelo estimado.

O exemplo mais simples de regressão linear consiste em ajustar uma reta a um conjunto de pares de observação em que  $y$  é a variável dependente,  $x_1$  é a variável explanatória ou regressora,  $e$  o termo de erro e  $i$ , o indicador da  $i$ -ésima observação. A expressão matemática desse modelo de regressão é dada por

$$y_i = f(x_{1i}) = \beta_0 + \beta_1 x_{1i} + e_i \quad (3.1)$$

Onde  $\beta_0$  é o coeficiente que representa o intercepto com o eixo dos  $y$  e  $\beta_1$  representa inclinação que a reta forma com o eixo das abscissas. O termo  $e$  é o erro ou resíduo entre o valor ajustado pelo modelo de regressão e o valor observado.

Dizemos que o erro ou resíduo é a discrepância entre o valor verdadeiro e o valor aproximado de  $y$ , que pode ser expresso, após organizar a equação (3.1), por:

$$e_i = y_i - (\beta_0 + \beta_1 x_{1i}) \quad (3.2)$$

Nesse caso, considerando o método dos mínimos quadrados ordinários, o melhor ajuste será aquele que minimizará a soma dos quadrados dos resíduos entre o valor de  $y$  medido e o valor de  $y$  calculado por meio do modelo de regressão linear, para todos os dados disponíveis na amostra, definida pela equação (3.3) a seguir:

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i})^2 \quad (3.3)$$

Como já visto, em engenharia de avaliações o mais comum é a utilização do modelo de regressão linear múltipla. Nesse sentido, uma extensão bastante útil da regressão linear simples é o caso no qual  $y$  é uma função linear de duas ou mais variáveis independentes  $x_1, x_2, \dots, x_m$ . Nesse caso, o modelo seria dado por (3.4):

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{mi}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + e_i \quad (3.4)$$

Esse modelo é, particularmente, útil quando se deseja fazer o ajuste de dados experimentais em que a variável explicada é uma função de duas ou mais variáveis explicativas. Nesse caso, a reta de regressão se torna um plano ou hiperplano, dependendo da quantidade de variáveis explicativas utilizadas no modelo.

Nesse caso, o erro ou resíduo é dado pela reorganização da equação (3.4):

$$e_i = y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}) \quad (3.5)$$

Assim como no caso anterior, o melhor ajuste é aquele em que os valores dos coeficientes são determinados escrevendo-se a soma dos quadrados dos resíduos, e derivando-a com relação a cada um dos coeficientes desconhecidos, onde se tem:

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \dots - \beta_m x_{mi})^2 \quad (3.6)$$

Dessa forma, os coeficientes  $\beta_0, \beta_1, \dots, \beta_m$  do modelo de regressão linear múltipla, fornecendo a soma mínima dos quadrados dos resíduos para o melhor ajuste, são obtidos igualando-se as derivadas parciais à zero.

Os modelos (3.1) e (3.4) podem ser obtidos pelo critério dos mínimos quadrados em que se realiza a regressão linear simples e regressão linear múltipla para se ajustar uma reta ou um plano de  $m$  dimensões aos  $n$  dados da amostra.

Segundo Schiavo *et al.* (2003), a análise de regressão múltipla em trabalhos de avaliação se consagrou a partir de meados da década de 80, estimulada pelo aparecimento dos computadores e softwares aplicativos específicos da área.

Com relação a esse avanço nos trabalhos de avaliação decorrentes do uso de modelos de regressão linear múltipla, essa abordagem econometrística estendeu o uso da análise de regressão para determinar o valor de um imóvel em função de seus vários atributos estruturais, como área total, número de dormitórios, padrão construtivo e número de vagas, de seus atributos de localização, como acessibilidade e vizinhança, bem como de seus atributos relacionados às condições de transação, como forma de pagamento e época da venda, os quais, também, podem afetar nos preços praticados pelo mercado imobiliário (UBERTI, 2016, p. 14).

Destaca-se que os coeficientes da regressão podem ser determinados por meio do método dos mínimos quadrados ordinários (OLS), ou da máxima verossimilhança (ML). A estimativa dos coeficientes é sem viés, e a variância calculada por meio do OLS ou ML é a menor dentre os estimadores não enviesados.

O uso do modelo clássico de regressão linear múltipla deve atender a alguns pressupostos básicos, os quais serão apresentados na sequência deste trabalho.

### 3.1.1 Inferência Estatística em Regressões Lineares

Em engenharia de avaliações, a inferência explica o comportamento do mercado que se analisa, com base em alguns dados levantados dele. Neste caso, a inferência estatística é fundamental para solucionar a questão, pois, conhecendo apenas uma parte do mercado, pode-se concluir sobre o seu comportamento geral, com determinado grau de confiança (DANTAS, 2012).

Conforme González (1997, p. 72), a aplicação da estatística inferencial possibilitou o surgimento de procedimentos de avaliação de valores de imóveis com maior precisão e com características científicas não encontradas nos métodos que empregam grande dose de subjetividade [...]. Todavia, de acordo com Thofehrn (2008, p. 79), a inferência estatística exige muita experiência em avaliação de imóveis, porque a ausência de variáveis importantes ou a inclusão de variáveis inadequadas pode conduzir a erros gravíssimos na avaliação do bem.

O objetivo da inferência, por meio da análise de regressão, é encontrar uma função linear que permita compreender a relação entre os elementos, além de estimar uma variável em função de uma ou mais variáveis (RADEGAZ, 2013).

Conforme Dantas (2012), na prática trabalha-se com modelos lineares ou linearizáveis, por facilidades no cálculo das estimativas das médias e facilidades de interpretação. Os modelos linearizáveis são aqueles que podem ser transformados em lineares pela simples transformação de suas variáveis.

Na maioria das situações, o engenheiro vai observar que são diversas as variáveis que influenciam na formação do valor de mercado de um imóvel. Assim, conforme Mendonça *et al.* (1998), o profissional deve identificar essas variáveis e encontrar o modelo explicativo do valor por meio das regressões múltiplas.

O modelo de regressão linear múltipla deve ser adotado quando mais de uma variável independente é necessária para explicar a variabilidade dos preços praticados no mercado. Conforme Dantas (2012), em engenharia de avaliações, geralmente, trabalha-se com modelos de regressão múltipla, tendo em vista a multiplicidade de fatores que interferem nos preços de um bem.

Sendo assim, como é inviável o levantamento de todos os dados de mercado de uma população, na prática, trabalha-se com um subconjunto de  $n$  elementos desta população, denominado amostra, por meio da qual, utilizando a inferência estatística, estimam-se os respectivos parâmetros populacionais.

A inferência estatística, por meio da regressão linear múltipla, tem como objetivo eleger um modelo matemático que represente o mercado imobiliário da melhor forma, considerando todas as características que interferem nesse valor. Essa ferramenta tem emprego obrigatório na busca de um melhor grau de precisão na determinação do valor do bem avaliado. Conforme Tiryaki e Andrade (2017) cabe, portanto, à estatística inferencial, o processo de generalização por meio de estimativa dos parâmetros populacionais e de testes de hipóteses estatísticas.

Segundo a NBR 14.653-2, na modelagem devem ser expostas as hipóteses relativas aos comportamentos das variáveis dependentes e independentes, com base no conhecimento que o engenheiro de avaliações tem a respeito do mercado, quando serão formuladas as hipóteses nulas e alternativas para cada parâmetro.

Para González (1997), o analista deve estipular modelos com as hipóteses de relacionamento entre as variáveis, que devem ser testadas pelos critérios estatísticos, verificando-se a validade destas hipóteses, ou seja, se os modelos são capazes de representar o segmento de mercado em questão. Para tanto, devem ser coletados dados de transações (evidências do mercado), analisando-se o ajuste dos modelos considerados a estes dados, dentro de um determinado grau de precisão. Os testes estatísticos permitem avaliar o próprio modelo e a importância individual das variáveis incluídas, indicando a qualidade geral do modelo formulado.

Durante a fase de modelagem, muitas etapas de análise são necessárias, tais como, a verificação de dados atípicos, se há variáveis que são pouco significativas, as tendências do modelo ou os pressupostos sobre os erros aleatórios, sendo que a variável resposta, no final do processo, deve ser explicada na escala original.

Nesse sentido, é conveniente destacar que o processo de análise de regressão exige que sejam atendidos alguns requisitos, chamados de pressupostos básicos do modelo, que devem ser verificados a sua não violação, assim que os parâmetros do modelo de regressão forem determinados. Em geral, os principais pressupostos recaem sobre a componente  $e$ , que reflete os resíduos do modelo. Dessa forma, o processo de análise de regressão exige o respeito aos pressupostos básicos, e ainda a outras condições relacionadas, que precisam ser atendidos para que a análise seja válida e possam ser realizadas inferências com o modelo obtido.

Segundo González (1997), para que os modelos de regressão obtidos sejam considerados aptos, deve-se garantir que: 1) Há homocedasticidade dos resíduos (a variância é constante); 2) Existe independência serial dos resíduos (não há

autocorrelação); 3) Os resíduos seguem a distribuição normal (normalidade); 4) A relação entre as variáveis independentes e a variável dependente é linear; 5) Não há colinearidade perfeita entre quaisquer variáveis independentes; além desses, o modelo deve ainda atender a outros requisitos, em parte decorrentes dos próprios pressupostos básicos: 6) As variáveis importantes foram incluídas (o modelo especificado é similar ao real); 7) Não existem observações espúrias (elementos claramente não adaptados ao modelo, chamados de outliers); 8) As variáveis independentes não são aleatórias (somente a variável dependente pode ser estocástica); 9) Os resíduos têm média nula; 10) O número de observações (tamanho da amostra) é maior que o de coeficientes a ser estimado.

Contudo, por ser um setor peculiar e em função da dificuldade de se obter dados, o mercado imobiliário carrega consigo características que, ao se verificar todas essas condições junto ao modelo obtido, há chances de se violar várias delas.

Conforme Lai (2011), a aplicação do modelo clássico de regressão linear múltipla para a avaliação de imóveis tem produzido, muitas vezes, sérios problemas, principalmente por causa da multicolinearidade nas variáveis independentes, da não-linearidade dos dados e da inclusão de dados considerados outliers.

O autor cita, ainda, que outro grave problema desse método é que ele envolve o julgamento humano quando há necessidade de realizar transformações nas variáveis do modelo ou quando se analisa o comportamento das variáveis.

Segundo González (2002), a ruptura de algum dos pressupostos não impede a utilização dessa técnica, apenas dificulta ou impede a generalização dos resultados. Se não estão presentes as condições básicas, a análise passa a ter um caráter determinístico, podendo ser vista como um ajustamento de curvas e os resultados são restritos ao conjunto específico de dados considerado. Nesse caso, se todas as características do bem avaliado estiverem contidas dentro dos intervalos respectivos, não há grandes prejuízos que impeçam o prosseguimento da análise, contudo, para avaliação em massa, trata-se de um problema sério em função da diversidade de características entre os avaliados.

Para González (2002), quase todas as condições podem ser razoavelmente garantidas, ou mesmo bem controladas, existindo técnicas estatísticas para corrigir eventuais problemas. Entretanto, para a autocorrelação e para a definição da forma funcional, as soluções paliativas não são suficientes para eliminar as dificuldades.

A autocorrelação decorre da influência sobre os resíduos ( $e_i$ ) de variáveis

formadoras de valor importantes, não consideradas na equação. Um modelo sem perturbações deve apresentar resíduos independentes, distribuídos de forma aproximadamente igual à distribuição normal (MENDONÇA *et al.*, 1998).

Segundo Wyatt (1996), a autocorrelação é o relacionamento entre elementos consecutivos, cuja forma mais comum é a autocorrelação serial. Em função das características espaciais do mercado imobiliário, os modelos de regressão podem sofrer com um tipo de autocorrelação denominado geralmente de autocorrelação espacial, principalmente em análises baseadas em recortes espaciais ou no tempo.

A dependência espacial tem certa similaridade com a dependência temporal com a diferença de que, em séries temporais, o elemento  $x$  influencia  $y$ , mas  $y$  não influencia  $x$ , ao passo que em séries espaciais, o elemento  $x$  influencia  $y$  e vice-versa, que é o que ocorre com o mercado imobiliário, principalmente quando se trata de avaliação em massa de imóveis, o que acaba por gerar autocorrelação espacial.

Conforme González (2002, p. 67), se a variável tem distribuição espacial, as variações nos dados ocorrem em todas as direções, o que dificulta a análise. A autocorrelação espacial é geral e está ligada a características peculiares dos dados, espacialmente distribuídos, que acabam por guardar uma maior relação com os imóveis da vizinhança próxima do que com os mais distantes.

A estatística espacial, segundo Krempf (2004), traz resultados diferentes daqueles obtidos pela estatística clássica, sendo os primeiros mais robustos por incorporarem a dimensão espacial, como, também, são necessárias as informações sobre a localização, e pressupõe-se que os dados são espacialmente dependentes.

Por fim, um outro problema ao se utilizar modelos clássicos de regressão linear é a forma funcional do modelo adotado, ou seja, das variáveis observadas, quais são importantes para que o modelo fique adequado e quais as transformações corretas a se fazer nas variáveis consideradas no modelo de regressão.

Para González (2002), não havendo indicações teóricas da forma de relacionamento das variáveis, a análise da função de regressão para verificar se ela é apropriada para os dados ou não, geralmente, é realizada por tentativas, por meio de gráficos de resíduos contra as variáveis dependentes ou independentes.

Caso haja uma forma definida nos resíduos, com tendências de crescimento ou curvaturas, tem-se um indicador favorável à rejeição da hipótese de variância constante para o erro e é possível que o modelo escolhido não seja o adequado.

### 3.2 MODELOS DE REGRESSÃO ESPACIAL

Em um estudo inovador para época, Tobler (1970, p. 3) afirmou que todas as coisas estão relacionadas entre si, entretanto coisas próximas estão mais relacionadas do que coisas distantes, expressão conhecida como 1<sup>a</sup> lei da geografia.

Sabe-se que a regressão clássica com utilização de inferência estatística utiliza em suas amostras de imóveis técnicas amplamente conhecidas, técnicas essas que acabam por, em alguns momentos, não considerar adequadamente o espaço geográfico como variável relevante para a formação dos valores dos imóveis.

Segundo Uberti (2016), quando a localização relativa entre os dados for uma variável de interesse, outros tipos de técnicas são necessárias para modelar estes fenômenos, as chamadas técnicas de estatística espacial ou de análise espacial.

Dessa forma, uma maneira de contornar a espacialidade dos dados é por meio dos modelos de regressão espacial. Neles, é tido como uma hipótese o fato de que o valor do imóvel pode estar correlacionado aos valores de outros imóveis, como pode estar correlacionado às características ou as amenidades vizinhas.

Conforme Serbanoiu (2019), na regressão linear múltipla, uma alteração de valor numa variável independente afeta apenas a sua variável dependente respectiva. Por exemplo, se na modelagem for estipulado que o valor venal de um apartamento depende, entre outros, do acabamento do prédio, uma variação na pontuação deste acabamento afeta apenas o valor venal daquele apartamento. Em alguns modelos de regressão espacial, uma variação na pontuação do acabamento de um prédio pode impactar no valor venal de apartamentos de prédios vizinhos.

Contudo, a utilização dos modelos tradicionais da economia e da econometria, quando se analisa fatores que tenham componente espacial, sofrem algumas limitações. Essa componente espacial possui características que não atendem às hipóteses básicas do modelo clássico de regressão linear. De modo geral, estas características estão divididas em heterogeneidade espacial e dependência ou autocorrelação espacial (ANSELIN, 1988).

Serrano e Valcarce (2000) concordam e acrescentam que, quando se trabalha com dados de natureza espacial, podem aparecer os denominados efeitos espaciais como a heterogeneidade e a autocorrelação espacial. A heterogeneidade aparece quando os dados utilizados para explicar um mesmo fenômeno são de unidades espaciais muito distintas, sendo que os problemas mais frequentes são a

instabilidade estrutural e a heterocedasticidade. A heterocedasticidade espacial ocorre pela omissão de variáveis ou outras formas de especificação que levam à aparição dos denominados erros de medidas. A dependência ou autocorrelação espacial surge sempre que o valor de uma variável em um lugar do espaço está relacionado com seu valor em outro ou outros lugares do espaço.

O conceito de heterogeneidade vem ao encontro do que propõe Anselin (1988), que diz que a heterogeneidade espacial está relacionada com a falta de estabilidade do comportamento ao longo do espaço e não é um problema, pois pode ser resolvida, na maioria dos casos, a partir de técnicas tradicionais de econometria.

Dantas (2003) completa dizendo que existem dois tipos de efeitos que podem ser encontrados nos dados distribuídos espacialmente: o efeito de heterogeneidade espacial e o efeito de autocorrelação ou dependência espacial. O primeiro, diz respeito à instabilidade dos parâmetros em relação à região em que se situam os dados e, na ausência de dependência espacial, podem ser diagnosticados pelo Teste de Chow e tratados pela metodologia tradicional, utilizando-se variáveis de interação entre cada variável explicativa e variável indicativa da região; o segundo, diz respeito a uma verdadeira interação espacial entre os dados, que pode afetar o termo de erro, a variável dependente ou ambos. Neste caso, somente a econometria espacial é capaz de realizar estimações seguras dos parâmetros do modelo.

O autor cita, ainda, que a autocorrelação espacial no termo de erro está relacionada a erros de medida ocasionados pelas divisões artificiais das unidades geográficas, como os limites estabelecidos para os bairros de uma cidade, que não necessariamente coincidem com a verdadeira dimensão do fenômeno observado. Em geral, no mercado imobiliário, há uma tendência de efeito de transbordamento de um bairro de maior importância sobre os bairros vizinhos.

Nesse viés, Anselin (1988) desenvolveu uma metodologia que utiliza técnicas econométricas para estudar a existência da dependência espacial. Essa dependência ocorre quando as observações de um local  $i$  dependem de outras observações situadas em locais  $j$ , sendo que  $i \neq j$ , e são modeladas de duas formas.

Segundo Anselin (2005), a dependência espacial pode ser incorporada nos modelos clássicos de regressão de duas formas: como regressor adicional na forma de uma variável dependente espacialmente defasada, representada por  $Wy$ , ou uma estrutura espacialmente defasada no erro da regressão, representada por  $Wu$ .

O primeiro modelo é conhecido como modelo de defasagem espacial (*Spatial Lag*) ou da variável dependente defasada e o segundo é conhecido como modelo do erro espacial (*Spatial Error*) ou do erro espacialmente correlacionado.

Os modelos de regressão espacial são vários, dependendo do tipo de autocorrelação que está presente. Segundo Arbia (2014, p.51), os modelos de regressão espacial têm a representação genérica mostrada na equação (3.7):

$$y = \rho \cdot Wy + Z\beta_1 + WZ\beta_2 + u \quad \text{com} \quad u = \lambda \cdot Wu + \varepsilon \quad (3.7)$$

Nesse caso, as componentes do vetor  $\varepsilon$  são independentes e identicamente distribuídas, e o vetor  $\varepsilon$  é independente do vetor  $Z$  (matriz de variáveis independentes). Impõe-se, também, que  $\varepsilon$  segue distribuição normal, de média zero e matriz de covariância constante e diagonal. Assim, a equação (3.7) é reescrita na forma (3.8):

$$y = \rho \cdot Wy + X\beta + u \quad \text{com} \quad u = \lambda \cdot Wu + \varepsilon \quad (3.8)$$

Onde,  $y$  é o vetor coluna de variáveis dependentes;

$X$  é a matriz de  $n \times k$  variáveis independentes;

$W$  é a matriz quadrada  $n \times n$  de pesos espaciais contendo relações de contiguidade de primeira ordem, padronizadas por linha ou funções distância;

$\rho$  é o coeficiente de autocorrelação espacial do termo  $Wy$ ;

$\lambda$  é o coeficiente de autocorrelação espacial do termo de erro  $Wu$ ;

$\beta$  é o vetor coluna de coeficientes estruturais, inclusive o termo de intercepto.

Esse é o modelo geral, pois considera tanto a possibilidade de autocorrelação espacial com uso do modelo de defasagem espacial ou da variável dependente defasada, quanto as possibilidades de autocorrelação espacial nos termos de erros com o modelo do erro espacial ou do erro espacialmente correlacionado.

Completando, Câmara *et al.* (2002) diz que a inclusão de efeitos espaciais em modelos de regressão pode ocorrer de diferentes formas, destacando-se:

**A)** Os modelos com efeitos espaciais globais: partem do princípio de que o processo espacial subjacente aos dados analisados é estacionário. Neste grupo estão o modelo da defasagem espacial (*Spatial Lag*) e o modelo do erro espacial (*Spatial Error*), cujos modelos serão estudados nesse tópico.

**B)** Os modelos com efeitos espaciais locais: ocorrem quando o processo espacial não é estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial, sendo que uma das alternativas é modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço. Nesse grupo temos a regressão geograficamente ponderada (GWR) que será discutida de forma pormenorizada no capítulo seguinte desse trabalho.

Simplificadamente, um processo estacionário tem a propriedade de que a média, variância e estrutura de autocorrelação não mudam no decorrer do tempo. O método de regressão espacial é o mais amplamente utilizado para obter modelos econôméticos, enquanto o método da *krigagem* é usado em geoestatística.

Conforme Dubin (2003), no primeiro método, a abordagem de matriz de peso usa uma matriz de conectividade para modelar as relações espaciais, enquanto, no segundo método, a abordagem geoestatística postula que a correlação entre observações é função da distância que separa suas localizações.

Nesse ponto, convém destacar o que afirma Trivelloni (2005, p. 34), ao citar que na literatura de avaliação imobiliária existem duas grandes escolas para o tratamento espacial dos dados de mercado: por um lado, alguns autores têm proposto a utilização das técnicas de econometria espacial e seus modelos de regressão espacial, enquanto outros têm proposto o uso de técnicas geoestatísticas. A discussão sobre qual das duas escolas têm fundamentos teóricos e metodológicos mais adequados para modelagem dos dados imobiliários continua em aberto.

Verifica-se que os dois métodos foram utilizados com certo sucesso na literatura, no entanto, há ainda pouca justificativa teórica na escolha do modelo e, por isso, é possível que o pesquisador estime um modelo mal especificado.

### 3.2.1 Modelo da Defasagem Espacial (*Spatial Lag*)

No modelo de defasagem espacial, conhecido em inglês por SAR (*Spatial Lag Model* ou *Spatial Auto Regressive Model*), a autocorrelação espacial é atribuída à variável explicada  $y$ . Nesse caso, quando  $Wu = 0$  em (3.8), temos o modelo SAR:

$$y = \rho \cdot Wy + X\beta + u \quad \text{com} \quad u = \varepsilon \quad (3.9)$$

Onde  $y$  é o vetor da variável dependente,  $\rho$  é o coeficiente de autocorrelação espacial que representa a influência média da unidade vizinha,  $W$  a matriz de vizinhança espacial ou matriz de ponderação espacial,  $X$  é a matriz das observações nas variáveis independentes dos dados,  $\beta$  é o vetor de parâmetros do modelo e  $\varepsilon$  é o vetor de erros aleatórios do modelo com média zero e variância constante.

Conforme pontua Almeida (2012), trata-se de um modelo de alcance global, caracterizado por hospedar a dependência espacial cujo alcance do transbordamento é global pela interferência do multiplicador espacial que faz com que o impacto sobre a variável dependente seja refletido para todas as regiões de estudo.

Destaca-se que nesse modelo não há autocorrelação no erro. A variável explicada depende das suas vizinhas, das variáveis explicativas e suas defasagens. Esse modelo ocorre, no contexto da engenharia de avaliações, quando o valor do imóvel é influenciado pelo valor das transações realizadas na vizinhança.

### 3.2.2 Modelo do Erro Espacial (*Spatial Error*)

No modelo do erro espacial, conhecido em inglês por CAR (*Spatial Error Model* ou *Conditional Auto Regressive Model*), a autocorrelação espacial é atribuída ao termo de erro  $u$ . Nesse modelo, os erros associados com qualquer observação são uma média dos erros das regiões vizinhas mais uma componente de erro aleatório. Nesse caso, quando  $Wy = 0$  em (3.8), temos o modelo CAR:

$$y = X\beta + u \quad \text{com} \quad u = \lambda \cdot Wu + \varepsilon \quad (3.10)$$

Onde  $y$  é o vetor da variável dependente,  $\lambda$  é o coeficiente de autocorrelação espacial que acompanha a defasagem  $Wu$ ,  $W$  a matriz de vizinhança espacial ou matriz de ponderação espacial,  $X$  é a matriz das observações nas variáveis independentes dos dados,  $\beta$  é o vetor de parâmetros do modelo e  $\varepsilon$  é o vetor de erros aleatórios do modelo com média zero e variância constante.

Conforme esclarece Almeida (2012), caracteriza-se também por ser um modelo de alcance global. Seu significado intuitivo é que o padrão espacial manifestado no termo do erro é dado por efeitos não modelados por conta da falta de adequada medida que, por sua vez, não são distribuídos aleatoriamente no espaço, mas, ao contrário, estão espacialmente autocorrelacionadas.

Nesse modelo deve-se supor a inexistência de autocorrelação na variável explicada. No contexto da engenharia de avaliações, isso resulta em valores venais não correlacionados entre si, fazendo, então, com que haja apenas a correlação no termo de erro  $u$ . Esse modelo ocorre, portanto, quando o termo do erro de um local está correlacionado com os valores do erro de outros locais da vizinhança.

### **3.2.3 Estimação do Modelo de Regressão Espacial**

Os modelos de regressão espacial podem ser estimados, conforme aponta Anselin (1999), utilizando vários métodos. O mais comumente usado é o de máxima verossimilhança, mas outros métodos também têm sido propostos, como os de variáveis instrumentais, mínimos quadrados espaciais, método dos momentos, método dos códigos e métodos bayesianos, entre outros.

Segundo Dantas *et al.* (2001), os dois modelos de regressão espacial não podem ser estimados diretamente pelo método dos mínimos quadrados ordinários, pois os estimadores assim gerados são viesados e inconsistentes, dado que existe viés de simultaneidade no modelo. O método de estimação por máxima verossimilhança é o mais indicado para estimar estes dois modelos de regressão.

### **3.2.4 Testes de Especificação**

Para Almeida (2012), quando se utiliza modelos econométricos espaciais o objetivo é remover a dependência espacial presente nos resíduos do modelo. Essa tarefa é realizada por meio da introdução de variáveis defasadas espacialmente do lado direito da regressão. Essas podem aparecer, principalmente, na forma de defasagem espacial na variável dependente ( $Wy$ ); ou como defasagem no termo do erro ( $Wu$ ), podendo ocorrer de forma isolada ou em conjunto.

A etapa de especificação consiste na identificação de qual tipo de variável defasada, espacialmente, deverá ser introduzida ao modelo econômico espacial. Os testes comumente utilizados para detectar a autocorrelação espacial são os testes I de Moran, o Multiplicador de Lagrange (LM) para defasagem espacial e, para o erro espacial, e o Multiplicador de Lagrange Robusto para ambos os casos.

Conforme Dantas (2003), o Teste I de Moran não identifica se a correlação espacial é devido ao erro ou a defasagem espacial, fazendo com que os testes dos

Multiplicadores de Lagrange sejam mais utilizados para identificar a autocorrelação espacial. Segundo Tyszler (2006), a vantagem do teste dos Multiplicadores de Lagrange é a exigência de se ter apenas os resíduos e os parâmetros de uma regressão pelo método dos mínimos quadrados ordinários.

### 3.2.4.1 Teste I de Índice Global de Moran

O índice global de Moran é um teste de autocorrelação amplamente conhecido e possibilita determinar se os resíduos do modelo de regressão pelo método dos mínimos quadrados ordinários apresentam autocorrelação espacial, verificando se há a presença da componente espacial no erro. O valor da estatística pode ser calculado pela expressão mostrada a seguir:

$$I = \frac{n}{S} \cdot \frac{e' We}{e'e} \quad (3.11)$$

Onde:  $e$  é o vetor de resíduos obtido por mínimos quadrados (variável padronizada);  
 $We$  é a matriz que representa os valores médios da variável de interesse padronizada nos vizinhos, definidos pela matriz de ponderação espacial  $W$ ;  
 $n$  é o número de observações ou regiões;  
 $S$  é a soma de todos os elementos da matriz de pesos espaciais  $W$ .

Conforme Vieira (2009), o índice global de Moran ( $I$ ) é uma das formas mais amplamente utilizadas de se medir a autocorrelação espacial. Essa estatística varia entre  $-1$  e  $1$ , fornecendo uma medida geral da associação linear (espacial) entre os vetores no tempo  $t$  e a média ponderada dos valores da vizinhança, ou lags espaciais. Valores próximos de zero indicam inexistência de autocorrelação espacial significativa, quanto mais próximo do valor unitário, mais autocorrelacionado estará. Se o valor dessa estatística for positivo, a autocorrelação será positiva, caso o valor dessa estatística seja negativo, a autocorrelação será negativa.

A estatística  $I$  de Moran tem aspectos positivos e negativos, segundo Tyszler (2006), evidenciados em alguns estudos. Tais estudos mostram que o teste de Moran apresenta um desempenho melhor que outros indicadores em termos de poder do teste para identificar correlações espaciais, porém não é capaz de diferenciar se a correlação espacial é nos erros ou na variável dependente.

### 3.2.4.2 Teste dos Multiplicadores de Lagrange

Os testes baseados no Multiplicador de Lagrange (LM) também são calculados a partir dos resíduos do modelo de mínimos quadrados, mas ao contrário do índice de Moran que não especifica o tipo de modelo espacial mais apropriado a testar, os testes LM estão orientados a determinar hipóteses nulas específicas para os modelos autorregressivos da variável dependente e do erro (TRIVELLONI, 2005).

Outros testes de Multiplicadores de Lagrange (LM) podem ser desenvolvidos, conforme Anselin (1999), para hipóteses mais complexas, como testes robustos à existência das duas formas de autocorrelação, processos autorregressivos de ordem superior, dados com variância no tempo e outros.

O teste dos multiplicadores de Lagrange robusto (LM robusto) para a defasagem identifica a presença de efeitos de defasagem espacial na variável dependente. Conforme Anselin (1996) e Dantas (2003), esse teste tem distribuição assintótica qui-quadrado com um grau de liberdade, sob a hipótese nula de não existência de defasagem espacial na variável dependente.

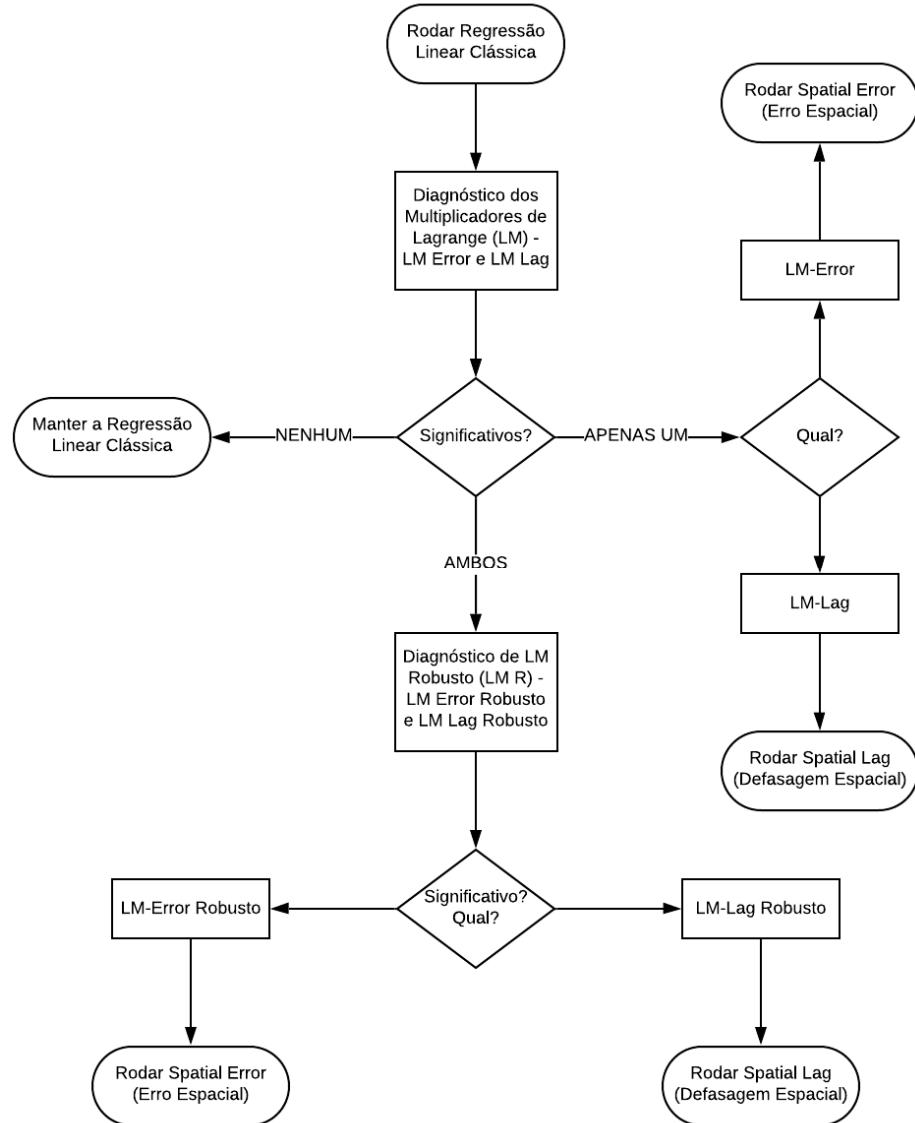
Desta forma, ao nível de significância  $\alpha$ , a hipótese de não-autocorrelação espacial na variável dependente do modelo clássico de regressão será rejeitada se a estatística do teste for superior ao ponto crítico da distribuição qui-quadrado.

Já o teste dos multiplicadores de Lagrange robusto (LM robusto) para o erro identifica efeitos de autocorrelação no termo do erro. Conforme Anselin (1996) e Dantas (2003), esse teste tem distribuição qui-quadrado com um grau de liberdade, sob a hipótese nula de não existência de defasagem espacial no termo de erro.

Nesse sentido, ao nível de significância  $\alpha$ , a hipótese de não-autocorrelação espacial nos resíduos do modelo clássico de regressão linear será rejeitada se a estatística do teste for superior ao ponto crítico da distribuição qui-quadrado.

Contudo, se ambos os testes dos LM forem significantes e se ambos LM robustos forem significantes, então deve-se observar qual teste LM robusto tem o menor  $p$ -valor ou escolher a alternativa com maior estatística LM robusto. É importante ressaltar que as versões robustas serão válidas somente se as versões padrões do teste forem significantes (ANSELIN, 1996 & HORBURG, 2015). O fluxograma da Figura 3 mostra a sequência a seguir para escolha do melhor modelo.

Figura 3: Fluxograma de decisão sobre o modelo adequado



Fonte: traduzido de Anselin (1996)

Almeida (2012) descreve o procedimento para a especificação dos modelos espaciais proposto por Anselin (1996). Este procedimento, mostrado na Figura 3, objetiva discriminar entre o modelo de defasagem espacial autorregressivo SAR e o modelo de defasagem do erro CAR, utilizando a combinação dos testes de ML tradicionais e ML nas suas versões robustas, como descrito nas etapas a seguir:

- Estima-se o modelo clássico de regressão linear (MCRL) pelo método MQO;
- Testa-se a hipótese de ausência de autocorrelação espacial devido a uma defasagem espacial ou a um erro espacial autorregressivo por meio do teste dos multiplicadores de Lagrange: ML para a defasagem e ML para o erro;

- C) Caso ambos os testes não sejam significativos do ponto de vista estatístico, estima-se o modelo clássico de regressão linear como o modelo mais apropriado para o estudo. Caso contrário, segue-se para o próximo passo;
- D) Se um deles for significativo, estima-se o modelo espacial indicado pela hipótese alternativa do teste. Por exemplo, se o teste ML para a defasagem for significativo, deve-se, inicialmente, estimar o modelo SAR para defasagem; e
- E) Caso ambos sejam significativos, estima-se o modelo apontado como o mais significativo pelas versões robustas desses testes ML robusto para a defasagem e ML robusto para o erro. Por exemplo, se ML robusto (defasagem) for significativo, escolhe-se o modelo SAR como o mais apropriado. Caso o ML robusto (erro) seja o significativo, adota-se o modelo CAR como o apropriado.

### 3.2.5 Matrizes de Ponderação Espacial ou Peso Espacial

Em geral, no estudo de econometria espacial, o formato mais comum para as matrizes de ponderação espacial é aquele construído a partir de características geográficas onde se atribui um peso mais elevado para as unidades geográficas mais próximas e um peso menos elevado para as unidades mais distantes.

Para isso deve ser definido o conceito de vizinhança de cada observação. Por meio de pesos que representam a covariância entre dois elementos, o operador espacial calcula para cada observação a média ponderada da variável aleatória na vizinhança, conforme aponta Uberti (2016). Para cada observação da variável aleatória  $y$  num ponto  $i$  o operador espacial  $Wy$  é obtido por meio da soma ponderada dos pesos  $W_{ij}$  de todos os valores de  $y$  nos pontos  $j$  da vizinhança:

$$Wy_i = \sum_{j=1}^n W_{ij}y_j \quad (3.12)$$

Sendo que os pesos  $W_{ij}$  não nulos definem a vizinhança de cada observação.

Por facilidade de interpretação, os pesos são padronizados por linha na matriz  $W$  de forma que a soma de pesos em cada linha é igual a 1. Consequentemente, a defasagem espacial da variável  $y$  no ponto  $i$  pode ser interpretada como sua média ponderada na vizinhança de  $i$  (UBERTI, 2016).

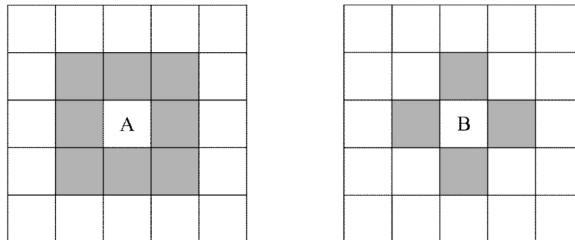
Almeida (2012) mostra que a matriz de ponderação ou pesos espaciais ( $W$ ), gerada a partir da proximidade geográfica, pode ser baseada na contiguidade ou em distância geográfica, conforme descrito nas etapas a seguir:

- **Contiguidade:** é baseada na ideia de que duas regiões vizinhas (contíguas) possuem maior interação espacial. Formalmente, essa matriz pode ser escrita por:

$$W_{ij} = \begin{cases} 1, & \text{se } i \text{ e } j \text{ são contíguos} \\ 0, & \text{se } i \text{ e } j \text{ não são contíguos} \end{cases} \quad (3.12)$$

Assim, os elementos  $W_{ij}$  têm valor zero na diagonal principal e para os imóveis não vizinhos. Já para os imóveis vizinhos é atribuído o valor um. A convenção de contiguidade é realizada de forma idêntica ao movimento das peças em um tabuleiro de xadrez. Assim, segundo Almeida (2012), as mais utilizadas são a convenção de contiguidade rainha (*queen*) e torre (*rook*) conforme a seguir:

Figura 4: Noções de Contiguidade de Rainha (A) e Torre (B)



Fonte: adaptado de Almeida (2012)

- **Distância Geográfica:** é baseada na ideia de que duas regiões próximas geograficamente têm uma maior interação espacial. Há dois formatos bastante conhecidos:

**1)** No primeiro caso, tem-se uma matriz  $W$  composta pelos  $k$  vizinhos mais próximos,  $w_{ij}(k)$ . Essa é uma matriz binária que pode ser formalmente escrita como:

$$W_{ij}(k) = \begin{cases} 1, & \text{se } d_{ij} \leq d_i(k) \\ 0, & \text{se } d_{ij} > d_i(k) \end{cases} \quad (3.13)$$

Onde  $d_i(k)$  é a distância de corte para a região  $i$  especificamente, a fim de que essa região  $i$  tenha exatamente  $k$  vizinhos, em que se assume  $w_{ii}(k) = 0$ .

**2)** No segundo caso, tem-se uma matriz  $W$  composta pelos  $k$  vizinhos mais próximos,  $w_{ij}(k)$ , contudo ela é baseada no inverso da distância e pode ser escrita como:

$$W_{ij}(k) = \begin{cases} 1/d_{ij}, & \text{se } d_{ij} \leq d_i(k) \\ 0, & \text{se } d_{ij} > d_i(k) \end{cases} \quad (3.14)$$

Onde  $d_i(k)$  é a distância de corte para a região  $i$ . Contudo, conforme Almeida (2012), agora, a influência de algum desses  $k$  vizinhos não é a mesma sobre a região  $i$ , sendo, aliás, diferente dependendo do inverso da distância que os separa.

Nota-se que Equações 3.13 e 3.14 poderiam ser reescritas para qualquer  $d_{ij}$ , independente do número de vizinhos. Neste caso, não existiria  $W_{ij} = 0$ , mas sim, valores que tenderiam a zero quando  $d_{ij}$  fosse suficientemente grande.

As matrizes de pesos espaciais mais tradicionais são construídas a partir de atributos físicos e geográficos, como vizinhança, distâncias geográficas e tempo de deslocamento. De acordo com a distância geográfica, a construção da matriz  $W$  está baseada no ordenamento de um espaço cartesiano representado por latitudes e longitudes. Esse tipo de ordenamento permite calcular as distâncias de quaisquer pontos no espaço. Com relação à dependência espacial, pressupõe-se que o grau de dependência é negativamente relacionado com a distância (VIEIRA, 2009).

Nesse caso, presume-se facilmente que a intensidade da dependência espacial aumenta quando a distância entre os dados diminui, e vice-versa.

Conforme Uberti (2016), em avaliações de imóveis as matrizes de pesos espaciais são construídas em função da distância geográfica e/ou número de vizinhos mais próximos. A matriz baseada na contiguidade não se aplica, pois, as amostras estão separadas geograficamente, não possuindo limites em comum.

Para matriz de vizinhos mais próximos, Almeida (2012) propõe que se faça:

- A) Roda-se o modelo clássico de regressão por mínimos quadrados ordinários;
- B) Testam-se os resíduos para autocorrelação espacial do teste I de Moran, usando  $L$  matrizes de  $k$  vizinhos mais próximos e  $L$  variando de  $k = 1$  a  $k = 20$ ; e
- C) Define-se  $k$  como o maior valor de I de Moran, significativo estatisticamente.

Por fim, é válido estimar os modelos de regressão, segundo Almeida (2012), usando mais de uma matriz de ponderação espacial e comparar os resultados, a fim de detectar discrepâncias significativas nos resultados, expondo a sensibilidade desses a escolha da matriz e checando, desta forma, a sua robustez.

## 4 REGRESSÃO GEOGRAFICAMENTE PONDERADA

Nos casos em que o padrão de autocorrelação espacial varia no espaço, o modelo de regressão espacial global não conseguirá explicar adequadamente a dependência espacial dos dados. Nesses casos, é indicado que se utilize um modelo de regressão local em que os parâmetros variem no espaço.

A técnica de regressão geograficamente ponderada (RGP), em inglês *geographically weighted regression* (GWR), foi inicialmente proposta por Brunsdon, Fotheringham e Charlton (1996) como um método para modelar a não estacionaridade (heterogeneidade) espacial, condição na qual a adoção de um modelo de regressão espacial global não conseguiria explicar adequadamente as relações entre algumas variáveis definidas em determinada região geográfica.

Para Carvalho *et al.* (2006), a ideia básica da GWR é ajustar um modelo de regressão para cada ponto no conjunto de dados, ponderando as observações por uma função de distância a este ponto. Isto corresponde a considerar que pontos mais próximos ao ponto em estudo tenham maior influência nos parâmetros estimados da regressão do que observações obtidas em pontos mais distantes.

Isso implica em se obter um conjunto de parâmetros ajustados para cada um dos pontos da região geográfica considerada. Conforme Carvalho *et al.* (2006), a proposta inicial tinha como objetivo fornecer uma ferramenta útil de análise exploratória por meio da identificação das variações entre variáveis ao longo do espaço, o que corresponde à não estacionaridade espacial.

Contudo, por auxiliar na explicação e previsão de fenômenos geográficos, este tipo de modelo pode ser classificado como confirmatório, desde que, conforme discutido no trabalho de Loureiro *et al.* (2006), seja avaliada com o devido cuidado sua premissa básica de influência do comportamento de uma determinada unidade geográfica sobre as suas vizinhas, e vice-versa.

Os modelos com efeitos espaciais são utilizados quando o processo é não-estacionário, e por isso, há a necessidade de que os coeficientes da regressão reflitam essa heterogeneidade espacial (Câmara *et al.*, 2002).

Considerando os modelos de regressão espacial global e de regressão espacial local, Fotheringham *et al.* (2002) apresentam um resumo com as principais características que diferem esses dois tipos de modelo, em se tratando de dados espaciais, conforme Tabela 1 apresentada a seguir:

Tabela 1: Características dos Modelos Globais e Modelos Locais

Modelos Globais	Modelos Locais
<ul style="list-style-type: none"> <li>- Resumem os dados para toda a região em estudo</li> <li>- Geram uma única estatística</li> <li>- Estatísticas não podem ser mapeadas ou analisadas por um SIG</li> <li>- Ênfase na similaridade no espaço</li> <li>- Procuram por regularidades ou padrões</li> </ul>	<ul style="list-style-type: none"> <li>- Desagregam localmente as estatísticas globais</li> <li>- Geram estatística para cada local</li> <li>- Estatísticas podem e devem ser mapeadas ou analisadas por um SIG</li> <li>- Ênfase nas diferenças no espaço</li> <li>- Procuram por exceções ou regiões em destaque</li> </ul>

Fonte: adaptada de Fotheringham *et al.* (2002)

Nos tópicos seguintes apresenta-se uma descrição mais detalhada do modelo de regressão geograficamente ponderada, bem como considerações sobre os diversos tipos de ponderações geográficas e as medidas de qualidade do ajuste.

#### 4.1 MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA

Pelo modelo clássico de regressão linear não-espacial em que a variável explicada  $y$  é representada como uma combinação linear de variáveis explicativas  $x_j$ , pode-se obter a estimativa de  $y$  para um ponto  $i$  a partir da equação de regressão:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j \cdot x_{ij} + \varepsilon_i \quad (4.1)$$

Onde,  $\varepsilon_i$  é o erro relacionado a  $i$ -ésima observação;

$y_i$  é a  $i$ -ésima observação da variável resposta, com  $i = 1, \dots, n$ ;

$x_{ij}$  é a  $j$ -ésima variável preditora para  $i$ -ésima observação, com  $j = 1, \dots, p$ .

Esse modelo assume que os coeficientes da relação linear entre a variável explicada e as variáveis explicativas são constantes, independente da posição geográfica. Essa relação é valida também para pontos situados em um mesmo *cluster*. Considerando o modelo de regressão linear múltipla, os coeficientes  $\beta_j$  são determinados globalmente pela resolução do problema de minimização a seguir:

$$\min_{\beta_0, \beta_1, \dots, \beta_j} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \cdot x_{i1} - \beta_2 \cdot x_{i2} - \dots - \beta_j \cdot x_{in})^2 \quad (4.2)$$

Segundo Carvalho *et al.* (2006), para se obter a estimativa dos coeficientes angulares e do intercepto em um modelo de regressão linear convencional, normalmente utiliza-se o método dos mínimos quadrados simples, resultando equação dos estimadores em notação matricial, mostrada em (4.3):

$$\hat{\beta} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{Y} \quad (4.3)$$

Onde,  $\hat{\beta}$  é o vetor dos coeficientes (intercepto e coeficientes angulares);

$\mathbf{X}$  é a matriz de observações das variáveis explicativas;

$\mathbf{Y}$  é o vetor de observações da variável dependente (explicada).

A regressão geograficamente ponderada (GWR) estende a estrutura da regressão clássica mostrada em (4.1), permitindo, conforme Silva (2019), que os parâmetros locais sejam melhor estimados do que os parâmetros globais.

Para Almeida (2012), o modelo de regressão geograficamente ponderada (GWR) gera uma sequência de regressões lineares, estimadas para cada região, usando sub amostras de dados, ponderadas pela distância.

A ideia de atribuir pesos às observações individuais a partir de um ponto focal, chamado de ponto de regressão ou ponto de calibragem, embute o conceito de que a importância relativa é decrescente com a distância desse ponto.

Então, segundo Fotheringham *et al.* (1997), a regressão geograficamente ponderada consiste numa técnica simples de extensão da regressão convencional permitindo que parâmetros locais, em vez de globais, sejam estimados.

Sendo assim, dado um modelo clássico de regressão linear não-espacial, a expressão equivalente para a regressão geograficamente ponderada (GWR) é:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) \cdot x_{ij} + \varepsilon_i \quad (4.4)$$

Onde,  $(u_i, v_i)$  é a posição coordenada do  $i$ -ésimo ponto no espaço;

$\beta_j(u_i, v_i)$  é o resultado da função contínua dada pelo ponto  $i$ .

No modelo,  $(y_i; x_{i1}, x_{i2}, \dots, x_{ip})$  são observações da variável resposta  $y$  de variáveis explicativas  $x_1, x_2, x_3, \dots, x_p$ , no local  $(u_i, v_i)$  da região geográfica estudada e  $\beta_j(u_i, v_i)$  ( $j = 1, 2, \dots, p$ ) são  $p$  funções de posições geográficas desconhecidas. Os  $\beta_j(u_i, v_i)$  são estimados localmente em cada posição  $(u_i, v_i)$  pelo procedimento dos mínimos quadrados ponderados, no qual são utilizados alguns pesos de decaimento em função da distância. Cada conjunto dos coeficientes estimados em  $n$  locais pode produzir um mapa de variação que pode fornecer informações importantes sobre a não estacionariedade da relação de regressão (MEI, 2004).

Como no modelo clássico de regressão, aqui o termo de erro aleatório também segue distribuição normal com média zero e variância constante.

O modelo possibilita, dessa forma, definir uma superfície contínua dos valores dos parâmetros que refletem a variabilidade espacial. Em caso de dependência espacial, espera-se que os parâmetros estimados de localidades próximas tenham magnitudes e sinais relativamente similares (SILVA, 2019).

Pode-se observar, a partir das Equações (4.1) e (4.4), que a regressão clássica é um caso particular de regressão geograficamente ponderada. Esta simplificação ocorre quando não há variação espacial nos parâmetros.

Segundo Carvalho *et al.* (2006), a principal diferença entre os modelos está na extensão e generalidade obtida para cada localização  $i$ , passando de observação para um modelo específico. Assumindo que os parâmetros exibam algum grau de consistência espacial, pode-se considerar um subconjunto de localizações próximas a que está sendo estudada como fornecedor de informações adicionais.

Na regressão geograficamente ponderada (GWR), supondo que os pesos no local  $(u_i, v_i)$  são  $w_j(u_i, v_i)$ ,  $j = 1, 2, \dots, n$ , então os parâmetros são determinados, para cada localização geográfica  $(u_i, v_i)$ , resolvendo o problema de minimização:

$$\min_{\beta_0(u_i, v_i), \dots, \beta_p(u_i, v_i)} \sum_{j=1}^n w_j(u_i, v_i) \cdot (y_j - \beta_0(u_i, v_i) - \beta_1(u_i, v_i) \cdot x_{j1} - \dots - \beta_p(u_i, v_i) \cdot x_{jp})^2 \quad (4.5)$$

Para o problema de minimização, consideram-se as matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  como:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \text{e} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (4.6)$$

Conforme Carvalho *et al.* (2006), no método dos mínimos quadrados ponderados, fatores de ponderação são aplicados em cada diferença ao quadrado antes da minimização, de forma que a imprecisão de alguns preditores receba maior penalidade que outros. Sendo assim, na forma matricial, o modelo seria dada por:

$$\hat{\beta} = (\mathbf{X}^T \cdot \mathbf{W} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{W} \cdot \mathbf{Y} \quad (4.7)$$

Na Equação (4.7), tem-se que  $\mathbf{W}$  é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação e os outros elementos são nulos. A matriz  $\mathbf{W}$  pode ser escrita como:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \quad (4.8)$$

É importante destacar que, em um modelo de regressão não espacial dado pela Equação (4.1), os pesos atribuídos a cada uma das observações são também constantes, fazendo com que a estimativa dos parâmetros seja a mesma para todo o espaço, obtendo o que chamamos de estimativa global.

Já na regressão geograficamente ponderada (GWR), os mínimos quadrados ponderados são utilizados, segundo Carvalho *et al.* (2006), para ponderação de cada observação de acordo com sua proximidade ao ponto  $i$ , de forma que, como no modelo representado pela Equação (4.4), os estimadores variem de acordo com a localização do ponto em estudo  $i$ , pela variação de  $\mathbf{W}$ .

Na forma matricial, os estimadores ponderados em  $(u_i, v_i)$  de um modelo de regressão geograficamente ponderada (GWR) podem ser escritos como:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \cdot \mathbf{W}(u_i, v_i) \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{W}(u_i, v_i) \cdot \mathbf{Y} \quad (4.8)$$

Onde, da mesma forma que em (4.7),  $\mathbf{W}(u_i, v_i)$  é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação com relação ao ponto de posição  $(u_i, v_i)$  no espaço, e os elementos restantes são nulos.

$$\mathbf{W}(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \quad (4.9)$$

Na Equação (4.9),  $\mathbf{W}(u_i, v_i)$  é uma matriz diagonal e diferente para cada ponto  $i$ , contendo os pesos  $w_{ij}$  em sua diagonal principal, obtidos por meio das funções de ponderação ou *kernel* (SILVA, 2019), que será apresentado a seguir.

Para Carvalho *et al.* (2006), uma abordagem geral, ao se tratar este tipo de modelo, é notar que, apesar da obtenção de estimativas não tendenciosas ser impossível, já que os coeficientes mostrarão um desvio ao longo do subconjunto local de calibração, pode-se ainda conseguir uma tendenciosidade pequena.

De fato, como Fotheringham *et al.* (1997) comentam, o processo de modelagem em regressão geograficamente ponderada pode ser visto como um equilíbrio de troca entre tendenciosidade e erro padrão.

Quanto maior for o subconjunto local de calibração, menores serão os erros padrões das estimativas dos coeficientes; no entanto, deve-se atentar para o fato de que quanto maior o subconjunto local, maior a probabilidade de que os desvios introduzidos pela região espacial acrescentem tendenciosidade (CARVALHO, 2006).

Importante destacar que os pressupostos básicos do modelo clássico de regressão linear permanecem para a regressão geograficamente ponderada (GWR).

## 4.2 PONDERAÇÃO ESPACIAL E PROCESSO DE CALIBRAÇÃO

Como visto, uma das formas de reduzir a tendenciosidade no modelo de regressão é considerar a função de ponderação espacial, de modo que posições geográficas mais afastadas do ponto  $i$  atuem de forma menos influente no momento da calibração do modelo. Nesse tópico será abordado a calibração e a ponderação espacial de modelos de regressão geograficamente ponderada (GWR).

### 4.2.1 Funções de Ponderação Espacial

O modelo de regressão geograficamente ponderada (GWR) considera uma ponderação geográfica na estimação dos parâmetros do modelo, onde cada observação da amostra é ponderada em função da sua distância para o ponto base

da regressão. Segundo Almeida (2012), as observações mais próximas possuem um peso maior, ao passo que as observações mais distantes, um peso menor.

Na regressão geograficamente ponderada (GWR),  $\mathbf{W}(u_i, v_i)$  é um vetor de ponderação baseado na proximidade de  $i$  aos locais de amostragem em torno de  $i$ .

Considerando inicialmente o modelo de regressão não-espacial mostrado em (4.1), tem-se que ele pode ser interpretado como um caso particular de (4.8) na qual:

$$w_{ij} = 1, \quad \forall i, j \quad (4.10)$$

Onde  $j$  representa um ponto específico no espaço no qual os dados são observados e  $i$  representa um ponto para o qual os parâmetros são estimados.

Nesse caso, em que cada observação tem um peso unitário, percebe-se uma estimativa constante dos parâmetros no espaço, o que significa que se está diante de um modelo de regressão espacial global, como visto anteriormente.

Conforme Fotheringham *et al.* (1998), um passo inicial em direção à ponderação com base na localização, seria excluir do processo de calibração local do modelo as observações que estão além de determinada distância  $d$  de influência. Isso seria equivalente a ajustar seus pesos a zero, o que corresponde à função:

$$w_{ij} = \begin{cases} 1, & \text{se } d_{ij} < d \\ 0, & \text{caso contrário} \end{cases} \quad \forall i, j \quad (4.11)$$

Em que  $d_{ij}$  representa a distância entre os pontos  $i$  e  $j$ . Fotheringham *et al.* (1998) afirma que o uso de (4.11) simplificaria o procedimento de calibração porque, para cada ponto em que os coeficientes devem ser calculados, apenas um subconjunto dos pontos de amostra precisa ser incluído no modelo de regressão.

Contudo, a função de ponderação espacial mostrada sofre o problema de descontinuidade. Fotheringham *et al.* (1998) observam que, à medida que  $i$  varia em torno da área de estudo, os coeficientes de regressão podem mudar drasticamente à medida que o ponto da amostra se move para dentro ou para fora da região circular em torno de  $i$ , ou seja, o modelo é fortemente dependente da distância  $d$ .

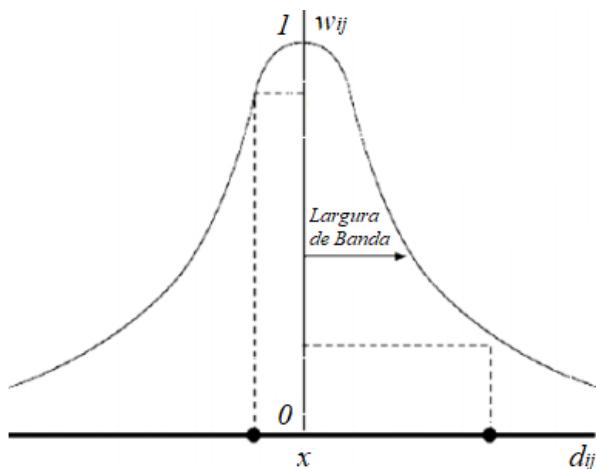
Uma maneira de evitar esse problema é especificar  $w_{ij}$  como uma função contínua de  $d_{ij}$ , que representa a distância entre  $i$  e  $j$ . Uma escolha possível seria:

$$w_{ij} = e^{-\frac{d_{ij}^2}{\beta^2}} \quad (4.12)$$

Em que  $\beta$  representa a largura de banda ou do inglês *bandwidth*, parâmetro que controla a variância da função de ponderação e determina a velocidade de decaimento do peso com a distância. Esse parâmetro de suavização pode ser fixo ou variar espacialmente de acordo com a disposição dos dados observados (SILVA, 2019). É fácil perceber que quando  $i$  e  $j$  coincidirem, a ponderação dos dados nesse ponto será a unidade. Nesse caso, a ponderação de outros dados diminuirá de acordo com uma curva gaussiana, conforme a distância entre  $i$  e  $j$  aumenta.

Segundo Uberti (2016), tal ponderação é feita pela função *kernel* espacial. De forma prática, o *kernel* usa a distância ( $d_{ij}$ ) entre dois pontos geográficos representando duas regiões, e um parâmetro de largura de banda ( $\beta$ ), para determinar um peso ( $w_{ij}$ ) entre essas duas regiões, que é inversamente relacionado à distância geográfica. A Figura 5 mostra como o *kernel* espacial opera.

Figura 5: Esquema de *Kernel Espacial* e *Bandwidth*



Fonte: adaptada de Fotheringham et al. (2002)

A Figura 5 ilustra o bandwidth em uma função de ponderação, onde  $x$  é o ponto de regressão, • é a observação,  $w_{ij}$  é o peso da observação  $j$  no ponto de regressão  $i$  e  $d_{ij}$  é a distância entre o ponto de regressão  $i$  e a observação  $j$ .

O uso de funções contínuas permite a inclusão de contribuições fracionárias fazendo com que pontos mais distantes tenham contribuições menos significativas.

Considerando a calibração de um modelo para o ponto  $i$ , por exemplo se  $w_{ij} = 0,5$ , os dados do ponto  $j$  contribuem apenas com metade do peso no processo de calibração. Para dados distantes de  $i$ , a ponderação cairá para praticamente zero, excluindo efetivamente essas observações da estimativa dos parâmetros do local  $i$ .

Para Carvalho *et al.* (2006), uma forma de conciliar as funções de ponderação, representadas nas Equações (4.11) e (4.12), é estimar  $\beta$  com base no valor crítico  $d$  de influência e um valor de tolerância  $e$ , conforme abaixo:

$$\beta = \frac{-\ln e}{d^2} \quad (4.13)$$

Onde  $e$  representa a proximidade a zero desejada. Um valor razoável seria  $e = 0,0001$  que corresponde a atribuição de peso virtualmente igual a zero para as observações do ponto  $j$  quando este dista linearmente  $d$  do ponto  $i$  em análise. O problema agora da calibração do modelo de regressão geograficamente ponderado reside na escolha adequada de valores para  $\beta$  ou, alternativamente, para  $d$ .

Na medida em que  $\beta$  tende a zero - ou, equivalentemente,  $d$  tende a maior distância entre  $i$  e  $j$  - os pesos tendem ao valor unitário para todos os pares  $(i, j)$  de pontos, de forma que os parâmetros estimados tornam-se uniformes, resultando em um modelo global. Por outro lado, à medida que  $\beta$  torna-se maior, as estimativas dos parâmetros dependerão cada vez mais das observações mais próximas de  $i$ , e, consequentemente, terão maior variância (CARVALHO *et al.*, 2006).

Para Uberti (2016), existem diversas formas de se definir a função *kernel* espacial a fim de se construir a matriz de ponderação espacial diagonal  $\mathbf{W}$ . Os pesos a serem obtidos pela função do *kernel* espacial podem ser binários, sendo a distância de varredura fixa, em um determinado valor, ou é possível especificar pesos que são uma função contínua da distância  $d_{ij}$  por meio de funções como a função *kernel* biquadrado de largura fixa, gaussiana fixa ou exponencial simples fixa.

Nesse caso, considerando que o número de observações dentro de uma subamostra pode variar consideravelmente, o *kernel* espacial de natureza fixa gera certa desvantagem, pois em algumas regiões, mais densas em dados, pode-se obter muitas observações, ao passo que em outras o conjunto de dados pode ser limitado.

Conforme Uberti (2016), para contornar esse problema foi definida uma função de ponderação adaptativa no sentido que se expanda em áreas em que as observações sejam escassas e encolha em áreas em que sejam abundantes.

Nesse caso, a função é chamada de adaptativa por ter a largura de banda expandida em áreas menos densas de dados com o objetivo de abrigar  $p$  vizinhos,

ao passo que em áreas mais densas, a largura de banda não precisa passar por tanta expansão para que consiga atingir os  $p$  vizinhos mais próximos.

Medina (2016) mostra que as duas principais funções de ponderações encontradas na literatura são as funções Gaussiana (em inglês *Gaussian*) e Biquadrática (em inglês *Bisquare*), conforme Tabela 2 abaixo adaptada do autor:

Tabela 2: Funções de Ponderação ou *Kernels*

Função de Ponderação	Equação da Função de Ponderação
(1) Gaussiana Fixa	$w_{ij} = \exp^{[-(d_{ij}/\beta)^2]}, \quad j = 1, 2, \dots, n.$
(2) Gaussiana Variável	$w_{ij} = \exp^{[-(d_{ij}/\beta_{i(k)})^2]}, \quad j = 1, 2, \dots, n.$
(3) Biquadrática Fixa	$w_{ij} = \begin{cases} [1 - (d_{ij}/\beta)^2]^2, & \text{se } d_{ij} \leq \beta \\ 0, & \text{se } d_{ij} > \beta \end{cases}, \quad j = 1, 2, \dots, n.$
(4) Biquadrática Variável	$w_{ij} = \begin{cases} [1 - (d_{ij}/\beta_{i(k)})^2]^2, & \text{se } d_{ij} \leq \beta \\ 0, & \text{se } d_{ij} > \beta \end{cases}, \quad j = 1, 2, \dots, n.$

Fonte: adaptada de Medina (2016)

Pela Tabela 2, pode-se notar que as principais funções são a Gaussiana e a Biquadrática e estas estão subdivididas em dois tipos: fixa ou variável, que diferem entre si pela escolha da largura de banda (*bandwidth*)  $\beta$ . Nas funções, o parâmetro  $d_{ij}$  refere-se a distância entre o ponto de regressão  $i$  e a observação  $j$ , o parâmetro  $\beta$  refere-se ao parâmetro de suavização (*bandwidth*) fixo, e o parâmetro  $\beta_{i(k)}$  refere-se a parâmetro de suavização (*bandwidth*) variável, com o índice  $k$  representando o número total de vizinhos mais próximos do ponto de regressão  $i$ .

A primeira (1) e terceira (3) funções de ponderação, chamadas de *Kernel* Gaussiana e *Kernel* Biquadrática, decrescem continuamente a medida que os pontos se distanciam e o seu parâmetro de suavização (*bandwidth*) é fixo. As funções desse tipo são chamadas de *kernel* espacial fixas, uma vez que assumem apenas um valor para o limite da largura de banda  $\beta$ , conforme visto na Figura 5.

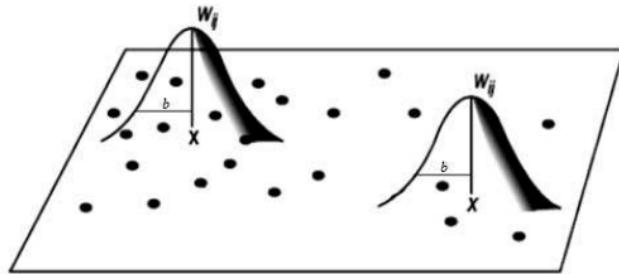
O parâmetro de suavização  $\beta$  tem o objetivo de controlar a variância da função de ponderação. Silva (2009) mostra que quando os dados são esparsos (espacialmente dispersos ou quando as áreas têm tamanhos diferentes), um mesmo

*bandwidth* pode ser adequado para algumas localidades e inadequados para outras pois, nesse último caso, os parâmetros estimados poderiam ter grandes erros padrões devido a poucos registros utilizados na estimação dos modelos.

Em algumas situações, como quando os dados não estão igualmente espaçados, é recomendável que o parâmetro de suavização da função de ponderação varie espacialmente de acordo com a disposição dos dados observados (SILVA, 2019). Desta forma, as regiões que apresentarem maior densidade de observações empregam uma função *kernel* que possua menor *bandwidth*, ao passo que as regiões com menor densidade de observações utilizam maior *bandwidth* (Figura 7). As funções (2) e (4) são exemplos de *kernel* que variam espacialmente.

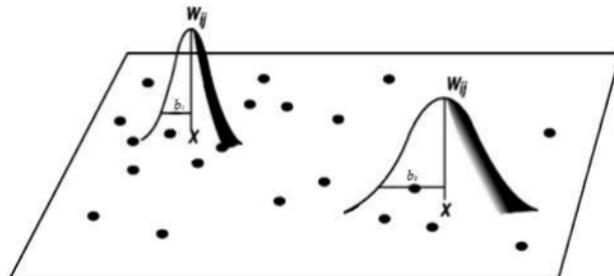
Nos casos em que as observações não são igualmente distribuídas entre as regiões analisadas, recomenda-se o uso de uma largura de banda variável. As Figuras 6 e 7 abaixo ilustram a utilização de largura de banda fixa e variável.

Figura 6: Funções de Ponderação Espacial com *Bandwidth* Fixo



Fonte: adaptada de Fotheringham et al. (2002)

Figura 7: Funções de Ponderação Espacial com *Bandwidth* Variável



Fonte: adaptada de Fotheringham et al. (2002)

Ao se utilizar uma largura de banda fixa para geração de um modelo via regressão geograficamente ponderada (GWR), o mesmo deverá ser especificado em unidades de distância, ao passo que, na utilização de uma largura de banda

variável, é necessário que se defina um número fixo de vizinhos  $k$ , mais próximos do ponto de regressão  $i$ , fazendo que, baseado no total de vizinhos  $k$ , o valor da largura de banda, expresso em unidades de distância, varie entre as regiões de estudo.

Na Tabela 3 pode-se verificar uma simulação que mostra os diferentes valores de peso  $w_{ij}$  obtidos para diferentes *bandwidths*,  $d_{ij}$  e funções de ponderação.

Tabela 3: Simulação de valores de pesos  $w_{ij}$

<i>Bandwidth</i> $\beta$ (em metros)	$d_{ij}$ (em metros)	$w_{ij}$	
		Gaussiana	Biquadrática
100	10	0,99005	0,98010
	50	0,77880	0,56250
	100	0,36788	0,00000
	500	0,00000	-
	1000	0,00000	-
500	10	0,99960	0,99920
	50	0,99005	0,98010
	100	0,96079	0,92160
	500	0,36788	0,00000
	1000	0,01832	-
1000	10	0,99990	0,99980
	50	0,99750	0,99501
	100	0,99005	0,98010
	500	0,77880	0,56250
	1000	0,36788	0,00000

Fonte: elaborada pelo autor

Pode-se observar que, quanto maior o valor da *bandwidth* e quanto menor o valor da distância  $d_{ij}$ , mais próximos são os resultados dos pesos obtidos pelos dois métodos (Gaussiano e Biquadrático). Contudo, à medida que a diferença entre *bandwidth* e  $d_{ij}$  vai reduzindo, o valor dos pesos  $w_{ij}$  obtidos pela função gaussiana passam a ser maiores que os valores dos pesos obtidos pela função biquadrática.

Observa-se ainda que, quando a distância entre  $i$  e  $j$  passa a ser maior que a *bandwidth*, a função gaussiana continua atribuindo pesos, ainda que pequenos, ao ajuste do modelo, ao passo que a função biquadrática atribui peso zero para as distâncias  $d_{ij}$  maiores ou iguais a *bandwidth* adotada. Esse comportamento da função gaussiana se deve principalmente ao fato de a curva normal ser assintótica em relação ao eixo das abscissas, fazendo com que os pesos não cheguem ao zero.

Conforme Fotheringham *et al.* (2002), para qualquer que seja a função de ponderação utilizada, a ideia central da regressão geograficamente ponderada é

que, para cada ponto  $i$ , existe um "aumento de influência" em torno de  $i$ , correspondente à função de ponderação, de modo que as observações próximas de  $i$  tenham mais influência na estimativa dos parâmetros de  $i$  do que as mais distantes.

Fotheringham *et al.* (2002) mostram ainda que os resultados da regressão geograficamente ponderada são relativamente insensíveis à escolha da função de ponderação, no entanto, são muito sensíveis à escolha do parâmetro de suavização.

#### 4.2.2 Processo de Calibração da Função de Ponderação Espacial

Segundo Fotheringham *et al.* (1998), uma dificuldade ao utilizar modelos de regressão geograficamente ponderada é que os parâmetros estimados são, em parte, funções do *kernel* de ponderação selecionado no método. Na Equação (4.11), por exemplo, à medida que  $d$  se torna maior, mais próxima será a solução do modelo pelo método dos mínimos quadrados ordinários, e quando  $d$  for igual à distância máxima entre os pontos no sistema, os dois modelos serão exatamente iguais.

Já quando se analisa a Equação (4.12), por exemplo, considerando a hipótese de que o valor do *bandwidth*  $\beta$  adotado tenda ao infinito, o peso  $w_{ij}$  tenderá a unidade para todos os pares de pontos, fazendo com que os parâmetros estimados se tornem uniformes e a regressão geograficamente ponderada se torne equivalente a regressão pelo método dos mínimos quadrados ordinários.

Em contrapartida, a medida que o valor do *bandwidth* vai se aproximando de zero, as estimativas dos parâmetros experimentam maiores variações pelo fato de dependerem cada vez mais da proximidade entre as observações e o ponto  $i$ .

Nesse caso, conforme Fotheringham *et al.* (1998), o problema consiste em como selecionar uma função de largura de banda ou decaimento apropriada.

Uma proposta possível para se encontrar o parâmetro de suavização ótimo seria escolher  $\beta$  por meio do método dos mínimos quadrados, ou seja, buscar um valor do parâmetro  $\beta$  que possa minimizar a Equação (4.14) mostrada a seguir:

$$\Delta(\beta) = \sum_{i=1}^n [y_i - \hat{y}_i(\beta)]^2 \quad (4.14)$$

Onde  $\hat{y}_i(\beta)$  é o valor estimado pelo modelo de regressão geograficamente ponderada (4.4) em que o  $\beta$  indica a suavização da função de ponderação espacial.

Na prática, a plotagem de  $\Delta(\beta)$  contra o parâmetro  $\beta$  fornecerá orientação sobre a seleção de um valor apropriado para o parâmetro ou ele poderá ser selecionado automaticamente por uma técnica de otimização (LEUNG *et al.*, 2000).

Contudo, Fotheringham *et al.* (1998) alertam que, ao se supor  $\beta$  muito pequeno a ponto de que as ponderações de todos os pontos, exceto o próprio  $i$ , sejam desprezíveis, então os valores ajustados nos pontos amostrados tenderão aos valores observados (reais), de modo que o valor da expressão (4.14) se torne zero.

Nessa linha, Silva (2019) mostra que essa técnica apresenta uma limitação, pois para essa forma o valor  $\beta$  ótimo é o valor tal que apenas o próprio elemento  $i$  permaneça na estimativa dos parâmetros no local  $i$ , ou seja, a soma se iguala a zero, gerando um modelo com um parâmetro para cada observação.

Isso mostra que esse critério de otimização pode apresentar problemas e deixar de ser útil, principalmente neste caso limitante onde os parâmetros do modelo não são definidos. Uma solução para esse problema é uma abordagem de validação cruzada (*Cross Validation - CV*) sugerida para regressão local por Cleveland (1979), em que o valor procurado é o que minimiza a soma dada pela Equação (4.15):

$$SQE = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(\beta)]^2 \quad (4.15)$$

Onde em que,  $SQE$  é a soma dos quadrados dos erros (*Cross Validation*),  $y_i$  é o valor observado da variável explicativa no ponto  $i$ ,  $\hat{y}_{\neq i}$  é o valor ajustado para o local  $i$  (de acordo com  $\beta$  ou  $d$ ) desconsiderando o próprio ponto  $i$  do processo de calibração. Desta forma, somente os pontos na vizinhança de  $i$  serão selecionados.

Essa abordagem apresenta a propriedade de combater o efeito de 'envolver o entorno', o que é indesejável, pois quando  $\beta$  se torna muito pequeno, o modelo é calibrado apenas em amostras próximas a  $i$  e não no próprio  $i$ .

Pode-se observar ainda que, se os valores  $i$  e  $j$  forem iguais, o peso  $w_{ij}$  será igual a unidade, contudo, quando os valores de  $i$  e  $j$  coincidirem, para a validação cruzada, deve-se assumir que  $w_{ij}$  é virtualmente igual a zero.

Conforme Carvalho *et al.* (2006), a escolha final destes parâmetros ( $\beta$  ou  $d$ ) geralmente será baseada na busca da minimização de  $SQE$  do fenômeno em estudo.

É importante ainda chamar a atenção que  $SQE$ , embora utilizado no modelo local de GWR, é um parâmetro global que, em função desta característica, servirá para comparar globalmente o nível de desempenho da GWR em relação aos modelos convencionais e espaciais globais. Além disso, a associação deste parâmetro com a variabilidade total da variável dependente  $y$  permite avaliar coeficientes globais de determinação ( $R^2$ ) que também servirão para comparações.

Almeida (2012) concorda com o fato de que, quando a largura de banda não é conhecida, ela pode ser determinada pela adoção da técnica otimizadora de validação cruzada (*Cross Validation - CV*), e ele apresenta, ainda, uma forma alternativa para determinar a melhor largura de banda, que é por meio da minimização do critério de informação de Akaike ( $AIC$ ), ou seja, valores menores no critério de informação de Akaike indicam largura de banda mais adequada para modelagem.

Como forma de obter estimativas ainda mais próximas do real, Charlton *et al.* (2009) sugerem uma validação cruzada usando o critério de informação de Akaike corrigido ( $AIC_C$ ), que pode ser utilizada, também como uma medida de ajuste para a comparação de modelos, como pode ser visto no Tópico 4.4 seguinte.

A regressão geograficamente ponderada (GWR) não é livre de críticas. Segundo Almeida (2012), a inferência estatística é um dos pontos fracos dessa abordagem, ofuscando a apuração dos erros padrão dos coeficientes locais e a realização dos testes estatísticos do tipo  $t$ , além da possibilidade de que ocorra elevada colinearidade entre os coeficientes locais.

Com relação aos valores dos erros padrão dos coeficientes locais, alguns softwares, como por exemplo o ArcGis e o GWR4 têm implementados nos relatórios de saída estes valores calculados para cada coeficiente de todas as amostras. O GWR4 gera ainda os valores de  $t$  calculados para cada regressor de todos os dados amostrais, contudo, ainda segundo Uberti (2016), analisar cada um destes erros padrão e  $t$  calculados não é uma tarefa simples e tampouco fácil.

#### 4.3 TESTES APLICADOS À REGRESSÃO GEOPONDERADA

Entre os testes aplicados à regressão geograficamente ponderada (GWR) pode-se destacar o teste de estacionariedade espacial dos parâmetros, proposto por Leung *et al.* (2000), e o teste de autocorrelação dos resíduos do modelo.

### 4.3.1 Teste de Estacionaridade dos Parâmetros

Na regressão linear múltipla, estamos interessados na nulidade ou não de cada parâmetro da regressão. No caso da regressão geograficamente ponderada (GWR), Leung *et al.* (2000, p. 13) desenvolveram um teste para determinar a não estacionariedade espacial dos parâmetros no modelo.

O método de regressão geograficamente ponderada (GWR) demonstrou ser uma técnica útil para detectar a não estacionariedade espacial. Segundo Leung *et al.* (2000), ao saber se os parâmetros em um modelo de regressão geograficamente ponderada variam ou não significativamente, pode-se obter não apenas uma visão maior dos dados e da estrutura na qual os dados são examinados, mas também algumas orientações valiosas para a decisão em ações futuras.

Os autores comentam ainda que, do ponto de vista estatístico, a pergunta a seguir deve ser testada estatisticamente: O conjunto de estimativas dos parâmetros locais  $\beta_{ij}$  ( $i = 1, 2, 3, \dots, n$ ) exibe variação espacial significativa na região de estudo?

Conforme Mei (2004), a técnica de regressão geograficamente ponderada fornece uma maneira viável de testar a relação de regressão linear global para dados espaciais. Isso equivale a testar as seguintes hipóteses:

$$H_0: \beta_j(u_i, v_i) = \beta_j, \quad j = 1, 2, 3, \dots, p.$$

$H_1$ : pelo menos um dos  $\beta_j(u_i, v_i)$  está variando com os locais.

Constrói-se a estatística do teste com base na soma dos quadrados dos resíduos obtidos, respectivamente, sob  $H_0$  e  $H_1$ . Em  $H_0$ , ajusta-se o modelo de regressão linear correspondente pela abordagem de mínimos quadrados ordinários e obtém-se a soma dos quadrados residuais, conforme Equação (4.16):

$$RSS(H_0) = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{4.16}$$

Onde  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ . Em  $H_1$ , o modelo de regressão de coeficiente de variação espacial mostrado na Equação (4.4) é ajustado pela técnica de regressão geograficamente ponderada e obtém-se a soma dos quadrados residuais por (4.17):

$$RSS(H_1) = \mathbf{Y}^T(\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})\mathbf{Y} \tag{4.17}$$

Segundo Mei (2004), a estatística de teste é então construída pela equação:

$$F = \frac{RSS(H_0) - RSS(H_1)}{RSS(H_1)} \quad (4.18)$$

#### 4.3.2 Teste de Autocorrelação dos Resíduos

Um dos pressupostos básicos do modelo clássico de regressão linear estabelece que não haja autocorrelação ou correlação serial nos termos de erros. Da mesma forma que no modelo clássico de regressão linear, a autocorrelação espacial nos termos de erro pode invalidar o pressuposto de homocedasticidade (variância constante nos termos de erro) e prejudicar os resultados da inferência estatística.

Segundo Dantas (2012), para se verificar a autocorrelação, um gráfico de resíduos versus valores estimados com pontos aleatórios é um forte indicador de distribuição aleatória de erros independentes, assim como o teste de Breusch-Pagan, entre outros, é utilizado para verificar a homocedasticidade.

Em relação à regressão espacial e à regressão geograficamente ponderada (GWR), há testes específicos de verificação de autocorrelação dos resíduos e, a partir deste, pode-se efetuar testes específicos de verificação de homocedasticidade.

Além das duas estatísticas amplamente conhecidas, o índice I de Moran e o índice C de Geary, utilizados para regressão global e regressão local, há também o coeficiente G de Ord e Getis, usado para explorar a autocorrelação espacial entre os resíduos da regressão geograficamente ponderada (GWR).

Conforme Serbanou (2019), o índice I de Moran original utiliza a diferença entre o valor de um ponto e sua média, e a correlaciona com a diferença entre o valor de outro ponto e sua respectiva média. Essa correlação entre dois pontos diferentes do espaço depende da estrutura de vizinhança, representada por uma matriz. A ideia da correlação espacial é saber se dois pontos semelhantes e próximos um do outro, segundo a estrutura de vizinhança, influenciam um ao outro.

O cálculo do índice I de Moran é realizado por meio da Equação (4.19) dada:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n \bar{w}_{ij} \hat{\epsilon}_i \hat{\epsilon}_j}{\sum_{i=1}^n \hat{\epsilon}_i^2} = \frac{\hat{\boldsymbol{\epsilon}}^T \mathbf{W} \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}}} \quad (4.19)$$

O cálculo do índice C de Geary, por sua vez, é definido pela Equação (4.20):

$$C = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (\hat{\epsilon}_i - \hat{\epsilon}_j)^2}{\sum_{i=1}^n \hat{\epsilon}_i^2} = \frac{\hat{\epsilon}^T (\mathbf{D} - 2\bar{\mathbf{W}})\hat{\epsilon}}{\hat{\epsilon}^T \hat{\epsilon}} \quad (4.20)$$

Onde  $\mathbf{D} = \text{diag}(\bar{w}_{1.} + \bar{w}_{.1}, \bar{w}_{2.} + \bar{w}_{.2}, \dots, \bar{w}_{n.} + \bar{w}_{.n})$ ,  $\bar{w}_{i.} = \sum_{j=1}^n w_{ij}$  e  $\bar{w}_{.i} = \sum_{i=1}^n w_{ji}$ .

Segundo Serbanoiu (2019), o problema da autocorrelação se relaciona com as variáveis relevantes omitidas. Suponhamos que o valor de determinado tipo de imóvel seja dependente de uma variável relevante que ainda não foi implementada na modelagem. Os resíduos retornados pela regressão múltipla terão uma média igual a zero; no entanto, o custo da omissão dessa variável relevante pode ser a heterocedasticidade ou então resíduos correlacionados.

#### 4.3.3 Identificação dos Coeficientes Constantes do Modelo

Considerando que certas variáveis explicativas que influenciam a variável resposta podem ser de natureza global, enquanto outras podem ser de natureza local, Brunsdon *et al.* (1999) propuseram um modelo misto de regressão geograficamente ponderada (MGWR) em que alguns coeficientes no modelo (4.4) são assumidos como constantes e os demais podem variar na região estudada. Após reordenar as variáveis explicativas, um modelo MGWR é especificado como:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^q \beta_j \cdot x_{ij} + \sum_{j=q+1}^p \beta_j(u_i, v_i) \cdot x_{ij} + \varepsilon_i \quad (4.21)$$

Nesse caso, ao se tomar  $x_{i1} = 1$  ou  $x_{i,q+1} = 1$  para qualquer  $i$ , o modelo de regressão poderá envolver o intercepto constante ou uma variação espacial.

Conforme Mei (2004), quando um modelo misto do tipo MGWR é aplicado para se analisar um conjunto de dados do mundo real, deve-se primeiro determinar quais são os coeficientes que podem ser mantidos fixos e quais não.

Desta forma, para um dado  $k$ , com  $k$  variando de 1 a  $p$ , inclusive, testar se o coeficiente  $\beta_k(u_i, v_i)$  da  $k$ -ésima variável explicativa  $x_k$  é constante na região geográfica de estudo, equivale a testar as seguintes hipóteses:

$$H_0: \beta_k(u_1, v_1) = \beta_k(u_2, v_2) = \beta_k(u_3, v_3) = \dots = \beta_k(u_n, v_n)$$

$H_1$ : nem todos os  $\beta_k(u_i, v_i)$  são iguais,  $i = 1, 2, 3, \dots, n$

Logo que os coeficientes constantes em um modelo misto de regressão geograficamente ponderada (MGWR) são identificados, pode-se determinar quais os coeficientes constantes e os de variação espacial são importantes para refletir a não estacionariedade espacial do modelo de regressão.

#### 4.4 MEDIDAS DE QUALIDADE DE AJUSTE DO MODELO

O quadrado do coeficiente de correlação de Pearson em uma regressão linear múltipla é chamado de coeficiente de determinação múltiplo ou simplesmente  $R^2$ . No caso de modelos de regressão linear múltipla, procura-se conhecer a proporção da variação de  $y$  que é explicada, conjuntamente, pelas variáveis  $x_1, x_2, \dots, x_n$ . É, portanto, uma medida de proporção da variabilidade em uma variável que é explicada pela variabilidade das outras variáveis. É pouco comum que tenhamos uma correlação perfeita na prática, ou seja,  $R^2 = 1$ , porque existem muitos fatores que determinam as relações entre as mais diversas variáveis na vida real.

Conforme Almeida (2012), algumas medidas como o coeficiente de determinação ( $R^2$ ) e o coeficiente de determinação ajustado ( $R^2_{ajus}$ ), são utilizadas para avaliar a qualidade de um ajuste. Porém, tais medidas não são as mais recomendadas como um indicador de ajuste de regressão na presença de erros autocorrelacionados e/ou heterocedásticos. Estas consequências são similares às que acontecem quando uma variável explicativa significativa é omitida do modelo.

Nesse caso, o método normalmente utilizado para seleção de modelos de regressão tem como base os valores de máxima verossimilhança dos diferentes modelos, ponderando pela diferença no número de parâmetros estimados.

Conforme Uberti (2016), nos modelos com estrutura de dependência espacial utilizam-se os critérios de informação onde a avaliação do ajuste é penalizada por uma função do número de parâmetros. É necessário ainda levar em conta o número de parâmetros independentes ao se incluir funções espaciais nos modelos onde, para cada nova variável no modelo, acrescenta-se um parâmetro.

Há critérios bastante difundidos na literatura que podem ser adotados para a seleção de modelos, sendo que estes levam em consideração a complexidade do

modelo no critério de seleção. Essencialmente, esses critérios penalizam a verossimilhança tendo como base o número de variáveis do modelo e, eventualmente, o tamanho da amostra. Essa penalização é realizada subtraindo-se do valor da verossimilhança uma determinada quantia que dependerá da complexidade do modelo, sendo que, quanto mais parâmetros, mais complexo.

Recomenda-se que, quando da utilização da estimação por máxima verossimilhança, seja feita a combinação do valor da função de verossimilhança  $f(x_n | \theta)$  com o critério de informação de Akaike ( $AIC$ ) e o critério de informação bayesiano ( $BIC$ ), proposto por Schwarz (1978), a fim de escolher o melhor modelo.

É frequente o uso do critério de informação de Akaike ( $AIC$ ) e o critério de informação bayesiano ( $BIC$ ) para a seleção de modelos, sendo que o melhor modelo será aquele que apresentar o menor valor para essas estatísticas. Nesse caso, o critério de informação de Akaike ( $AIC$ ) e o critério de informação bayesiano ( $BIC$ ) incorporaram em seus cálculos o valor da função de máxima verossimilhança.

Em suas pesquisas, Akaike (1974) mostrou que o viés do modelo é dado assintoticamente por  $p$ , em que  $p$  é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como dado pela Equação (4.22):

$$AIC = -2 \ln f(x_n | \theta) + 2p \quad (4.22)$$

O critério de informação de Akaike ( $AIC$ ) é um critério assintótico, desta forma Bozdogan (1987), mais tarde, propôs uma correção ao  $AIC$ , sendo ele (4.23):

$$AIC_C = -2 \ln f(x_n | \theta) + 2p + 2 \cdot \frac{p(p+1)}{n-p-1} \quad (4.23)$$

Onde  $AIC_C$  é critério de informação de Akaike corrigido para quando o número de observações for menor do que dez vezes o número de parâmetros;  $n$  é o número de observações,  $p$  é o número de parâmetros estimados pelo modelo.

Segundo Burnham *et al.* (2002), o critério de informação de Akaike ( $AIC$ ) não é uma prova sobre o modelo, no sentido de testar hipóteses, mas uma ferramenta para a seleção de modelos, não há significância e nem valor- $p$ . Dado um conjunto de dados e vários modelos correspondentes, pode-se classificá-los de acordo com o seu  $AIC$ , com aqueles tendo os menores valores de  $AIC$  como sendo os melhores.

Com relação ao critério de informação bayesiano (*BIC*), proposto por Schwarz (1978), sendo  $f(x_n | \theta)$  um modelo estatístico estimado por meio do método da máxima verossimilhança, então o critério de informação bayesiano é dado por:

$$BIC = -2 \ln f(x_n | \theta) + p \ln n \quad (4.24)$$

Onde  $f(x_n | \theta)$  é o modelo escolhido,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da respectiva amostra coletada.

## 5 MATERIAIS, METODOLOGIA E ÁREA DE ESTUDO

O objetivo desse capítulo é detalhar o método empregado e a área de abrangência escolhida para essa pesquisa, fazendo uma exposição dos procedimentos adotados à condução da coleta, do tratamento e da análise dos dados.

### 5.1 CLASSIFICAÇÃO DA PESQUISA

Segundo Fonseca (2002 *apud* GERHARDT e SILVEIRA, 2009), metodologia é o estudo da organização, dos caminhos a serem percorridos, para se realizar uma pesquisa ou um estudo. Nesse sentido, metodologia significa o estudo dos caminhos, dos instrumentos utilizados para fazer uma pesquisa científica.

Em geral, pode-se classificar uma pesquisa científica quando aos objetivos (fins) e quando aos procedimentos metodológicos (meios) adotados no estudo.

Segundo Gil (2017), com base nos objetivos traçados, é possível classificar as pesquisas em três grandes grupos: exploratórias, explicativas e descritivas.

Em linhas gerais, as pesquisas exploratórias têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses; ao passo que as pesquisas explicativas têm como preocupação central identificar os fatores que contribuem para a ocorrência dos fenômenos.

Nesse estudo, o foco será a **Pesquisa Descritiva**, que têm como objetivo a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis. Algumas pesquisas descritivas vão além da simples identificação da existência de relações entre variáveis, e pretendem determinar a natureza dessa relação. Nesse caso, tem-se uma pesquisa descritiva que se aproxima bastante da pesquisa explicativa (GIL, 2017).

Segundo o referido autor, são em grande número as pesquisas que podem ser classificadas como descritivas e a maioria das que são realizadas com objetivos profissionais provavelmente se enquadram nesta categoria.

Já para Vergara (2015), esse estudo é classificado como descritivo quanto aos fins, uma vez que se propõe descrever um fenômeno ou objeto com a intenção de revelar particularidades e detalhes não perceptíveis normalmente.

Para se desenvolver uma pesquisa, é indispensável também selecionar o método de estudo a se utilizar. Para Gil (2017), a classificação das pesquisas em

exploratórias, descritivas e explicativas é útil para o estabelecimento de seu marco teórico, ou seja, para possibilitar uma aproximação conceitual. Todavia, para analisar os fatos do ponto de vista empírico, para confrontar a visão teórica com os dados da realidade, torna-se necessário traçar um modelo conceitual e operativo da pesquisa.

Nesse sentido, podem ser definidos dois grandes grupos: aqueles que se valem das chamadas fontes de “papel” e aqueles cujos dados são fornecidos por pessoas. No primeiro grupo, estão a pesquisa bibliográfica e a pesquisa documental. No segundo, estão as pesquisas experimentais e o estudo de caso (GIL, 2017).

Quanto aos procedimentos metodológicos adotados (meios), a pesquisa se classifica, segundo Vergara (2015), como experimental, pois é investigação empírica na qual o pesquisador manipula e controla variáveis independentes e observa as variações que tal manipulação e controle produzem em variáveis dependentes.

De modo geral, o experimento representa o melhor exemplo de pesquisa científica. Essencialmente, a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto.

As pesquisas experimentais constituem o mais valioso procedimento disponível aos cientistas para testar hipóteses que estabelecem relações de causa e efeito entre as variáveis. Em virtude de suas possibilidades de controle, os experimentos oferecem garantia muito maior do que qualquer outro delineamento de que a variável independente causa efeitos na variável dependente (GIL, 2017).

Conforme Vergara (2015), a variável é um valor que pode ser dado por quantidade, qualidade, característica, magnitude, variando em cada caso individual. Variável independente é aquela que influencia, determina ou afeta a dependente. É conhecida, aparece antes, é o antecedente. Variável dependente é aquela que vai ser afetada pela independente, é descoberta, consequente. A pesquisa experimental permite observar e analisar um fenômeno, sob condições determinadas.

O esquema básico da experimentação pode ser assim descrito: seja Z o fenômeno estudado, que em condições não experimentais se apresenta perante os fatores A, B, C e D. A primeira prova consiste em controlar cada um desses fatores, anulando sua influência, para observar o que ocorre com os restantes (GIL, 2017).

Conforme o autor, a pesquisa experimental constitui o delineamento mais prestigiado nos meios científicos. Consiste essencialmente em determinar um objeto

de estudo, selecionar as variáveis capazes de influenciá-lo e definir as formas de controle e de observação dos efeitos que a variável produz no objeto.

Desta forma, este estudo pode ser classificado quanto aos objetivos traçados (ou fins) como uma pesquisa descritiva, e quanto aos procedimentos metodológicos adotados (ou meios) como uma pesquisa experimental.

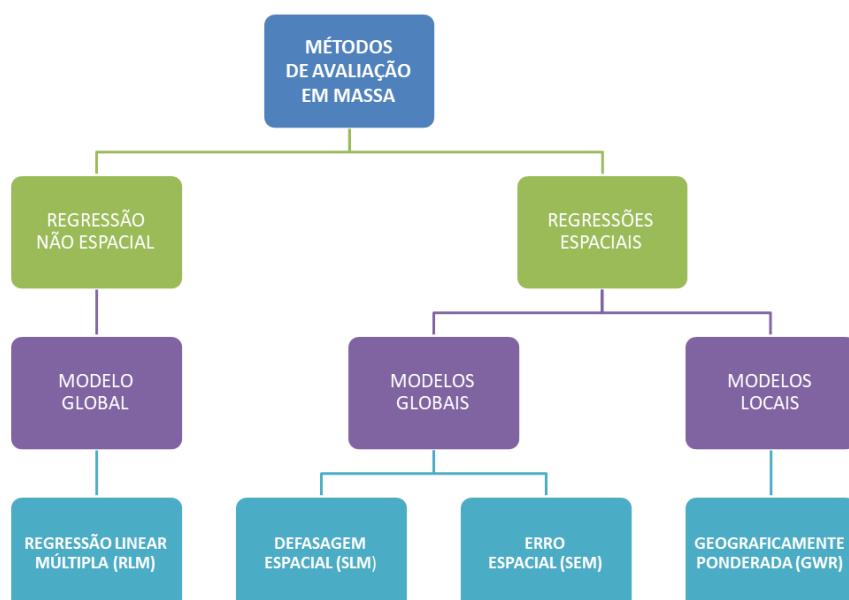
## 5.2 PROCEDIMENTOS METODOLÓGICOS

Buscou-se nesse estudo investigar o uso de regressão geograficamente ponderada (GWR) para a avaliação em massa de imóveis tipo apartamento. Para que o objetivo proposto fosse atingido, coletou-se 225 dados de mercado georreferenciados em sítios de imobiliárias, nos meses de março e abril de 2020, referente aos bairros Centro, Agronômica e Trindade, no município de Florianópolis/SC.

Destes dados, 35 foram retirados para compor a amostra de validação com o objetivo de analisar o desempenho das modelagens, e a outra parte foi destinada para elaboração dos modelos de regressão abordados nessa pesquisa.

Nesse sentido, utilizou-se os métodos de regressão linear clássica, de regressão espacial e de regressão geograficamente ponderada para a modelagem dos dados de mercado e para gerar uma superfície de valores representativa da área de pesquisa. O esquema abaixo contém os modelos que foram analisados:

Figura 8: Esquema contendo os modelos analisados



Fonte: o autor

Destaca-se que as modelagens globais serão realizadas com o objetivo de comparar a qualidade destas, por meio de indicadores estatísticos, com a qualidade da modelagem, objeto deste estudo: regressão geograficamente ponderada (GWR).

O valor de um imóvel depende de um conjunto de fatores entre os quais se destacam: as suas características físicas, o momento da avaliação, as condições legais a que está condicionado o imóvel e, na maioria das vezes, a localização.

Segundo Trivelloni (2005), para efeito de avaliação em massa de imóveis, quando se considera uma região geográfica de condições fiscais fixas, e uma amostra de mercado de imóveis em condições semelhantes de transação e coletada na mesma época, o valor dos imóveis poderá ser modelado pela equação a seguir:

$$V = f(L, E, C, T) \quad (5.1)$$

Onde,  $V$  representa o valor de mercado dos imóveis;

$L$  representa as variáveis de localização do imóvel, caso ela seja relevante;

$E$  representa fatores econômicos (oferta/transação);

$C$  representa as características físicas dos imóveis em função da tipologia;

$T$  representa a tipologia dos referidos imóveis.

Para verificar se a regressão geograficamente ponderada é uma alternativa viável e eficaz para a avaliação em massa de imóveis urbanos, realizou-se a coleta de dados de mercado nos bairros Centro, Agronômica e Trindade. Esta coleta foi realizada para se obter informações sobre os imóveis à venda, contendo a identificação do imóvel, infraestrutura da região, polos de valorização e desvalorização, características dos imóveis e seus respectivos valores.

Os dados de mercado foram tratados e submetidos para geração do melhor modelo de regressão clássica, do melhor modelo de regressão espacial e do melhor modelo de regressão geograficamente ponderada. Ressalta-se que por melhor modelo entende-se aquele que apresenta a melhor transformação, a melhor forma de representar a variável dependente e as variáveis independentes e que passou em todos os testes estatísticos necessários para o respectivo método.

Para melhor compreensão da descrição das etapas metodológicas realizadas a partir deste ponto, optou-se por separar em tópicos conforme segue:

### - ETAPA A: ANÁLISE EXPLORATÓRIA DOS DADOS DA AMOSTRA

Efetivada a coleta de dados, realizou-se a sua análise exploratória com o objetivo de reconhecê-los, fazer o tratamento e identificar como se dá a relação das variáveis explicativas entre si e destas com a variável explicada, valor unitário (VU).

Nessa etapa, com auxílio do *software R* 3.5.3, fez-se uma análise dos diagramas de dispersão das variáveis, a plotagem dos histogramas de frequências, a verificação de possíveis transformações nas variáveis por meio de diagramas *box-plot* e Box-Cox e os testes de significância e de correlação, com o objetivo de verificar variáveis relevantes para a formação dos valores dos imóveis em estudo.

Foram excluídas algumas variáveis que não se mostraram significativas para o estudo e seguiu-se com as variáveis restantes para a etapa seguinte.

### - ETAPA B: MODELAGEM DA REGRESSÃO LINEAR MÚLTIPLA

Nessa etapa buscou-se estimar a melhor equação de regressão linear múltipla pelo método dos mínimos quadrados ordinários, utilizando como ferramenta de modelagem o *software Infer32*. Foram testadas algumas transformações nas variáveis em estudo de modo que o modelo de regressão linear escolhido fosse o melhor modelo dentre os possíveis. Destaca-se que foram testados modelos com e sem as variáveis dicotômicas indicativas de bairro. Realizou-se, ainda, os testes com o objetivo de verificar os pressupostos básicos do modelo clássico de regressão.

### - ETAPA C: MODELAGEM DA REGRESSÃO DO ERRO ESPACIAL

Observou-se a existência de autocorrelação espacial nos resíduos. Nesse sentido, realizou-se novamente o tratamento dos dados utilizando-se agora o modelo de regressão espacial que melhor representava a realidade local. Fez-se, nessa etapa, a seleção da matriz de pesos espaciais por meio do resultado obtido na estatística I de Moran. Partiu-se, então, para escolha do melhor modelo, sendo analisados os desempenhos tanto do modelo do erro quanto da defesagem espacial, cuja escolha se deu pelo teste Multiplicador de Lagrange (LM). Nesta etapa utilizou-se os softwares *R* versão 3.5.3 para geração dos semivariogramas e *GeoDa* de Anselin (2005) para modelagem das regressões do erro e da defasagem espacial.

#### - ETAPA D: MODELAGEM DA REGRESSÃO GEOPONDERADA

Finalizada esta etapa, começou-se a modelagem dos dados por regressão geograficamente ponderada. Inicialmente, definiu-se que o critério de seleção do modelo seria aquele que apresentaria o menor valor no Critério de Informação de Akaike (AIC). Utilizou-se o software GWR4 criado por Fotheringham *et al.* (1998) para modelagem dos dados e testou-se, para todos os 4 tipos de função de ponderação (Gaussiano Fixo e Variável, Biquadrático Fixo e Variável), as mais diversas larguras de banda, por meio da função “*Golden Section Search*”, que busca a melhor largura de banda por meio de critérios definidos pelo usuário.

Fez-se, ainda, simulações objetivando ver se, dentre as variáveis explicativas consideradas no modelo, alguma tinha comportamento estacionário e que pudesse, então, ser tratada como variável global ao longo da área em estudo.

#### - ETAPA E: QUALIDADE E DESEMPENHO DOS MODELOS

Buscou-se, na sequência, verificar a qualidade das modelagens por meio de estatísticas conhecidas, como o Critério de Informação de Akaike normal (AIC) e corrigido (AICc), o Critério de Informação Bayesiano (BIC), o log Likelihood (LIK), o desvio padrão da regressão e o coeficiente de determinação ( $R^2$ ).

Fez-se ainda o gráfico de predição das modelagens e o cálculo da raiz quadrada do erro quadrático médio (RMSE) e do erro relativo das modelagens.

Para medir o desempenho das modelagens, realizou-se o cálculo dos indicadores propostos pela norma IAAO sobre a amostra de validação. Obteve-se, então, a correspondente mediana das razões de avaliação (Med R), o coeficiente de dispersão das razões de avaliação (COD) e o diferencial relativo ao preço (PRD).

#### - ETAPA F: CRIAÇÃO DAS SUPERFÍCIES DE VALORES

Para realização desta etapa definiu-se, inicialmente, o imóvel paradigma representativo da área em estudo, seguido pela homogeneização dos valores preditos por cada uma das modelagens. Destaca-se que, para regressão geograficamente ponderada (GWR), foi realizada a homogeneização de cada dado amostral com base na regressão obtida para o respectivo dado.

Realizada a homogeneização, buscou-se construir o semivariograma experimental e, por meio do *software* R na sua versão 3.5.3, calculou-se os parâmetros de alcance, efeito pepita e contribuição para cada um dos modelos teóricos: gaussiano, esférico e exponencial. Calculou-se, também, o RMSE e o  $R^2$  de cada modelo teórico gerado, por meio de um processo de validação cruzada.

Escolhido o semivariograma teórico que apresentava o menor RMSE e a mais forte dependência espacial e, portanto, melhor representava cada modelagem, procedeu-se com a construção das superfícies de valores unitários e dos erros.

Para etapa de construção das superfícies utilizou-se o *software* Surfer 15. Realizou-se, portanto, a krigagem ordinária dos valores homogeneizados preditos por cada modelo de regressão com base no semivariograma teórico escolhido. Construiu-se um mapa de predição dos valores unitários e dos erros da krigagem.

#### **- ETAPA G: QUALIDADE E DESEMPENHO DA KRIGAGEM**

Por fim, com o objetivo de avaliar a qualidade e o desempenho da krigagem ordinária, calculou-se, para a amostra de validação, o RMSE dos valores unitários preditos pela krigagem e, novamente, o valor dos parâmetros recomendados pela IAAO: mediana das razões de avaliação (Med R), o coeficiente de dispersão das razões de avaliação (COD) e o diferencial relativo ao preço (PRD).

### **5.3 CARACTERIZAÇÃO DA ÁREA DE ESTUDO**

Conhecido o método a ser utilizado para se alcançar os objetivos dessa pesquisa, pode-se delimitar a área de estudo. Desta forma, a aplicação do método proposto será realizada em três importantes bairros no distrito central da região de Florianópolis, em Santa Catarina, a saber: Centro, Agronômica e Trindade.

#### **5.3.1 Município de Florianópolis: Capital de Santa Catarina**

Conforme a Secretaria de Turismo (2019), Florianópolis é a capital do estado de Santa Catarina, conhecida também como Ilha da Magia. Situa-se no litoral catarinense, e conta com uma parte insular (ilha de Santa Catarina) e outra parte

continental incorporado à cidade em 1927, com a construção da ponte Hercílio Luz (Figura 9) - 820 m de comprimento - que ligou a ilha ao continente.

Figura 9: Ponte Hercílio Luz (Florianópolis)



Fonte: <[vemfloripar.com.br](http://vemfloripar.com.br)>. Acesso em 05 nov. 2019

O plano da cidade originou-se a partir da Praça XV de Novembro, que se estende até o pé da colina onde se eleva a catedral. A articulação dos bairros e subúrbios faz-se por meio de ruas ou avenidas, de longo e sinuoso traçado, entre o mar e as encostas dos morros. Apesar das sensíveis modificações, que construções modernas introduziram, a paisagem guarda ainda muito da arquitetura colonial.

Florianópolis é uma das três capitais insulares do Brasil. Vem se firmando cada vez mais como centro de turismo, graças às praias (Jurerê, Canasvieiras, Ingleses e outras) que circundam a ilha e à beleza da Lagoa da Conceição, a 13 Km de distância do centro. Nos arredores da lagoa, são características as rendas de bilros, de tradição açoriana. Entre os monumentos históricos da cidade, destacam-se a casa de Victor Meirelles, os fortões e a famosa catedral metropolitana.

Conforme a Secretaria de Turismo (2019), a maioria da população vive no continente e em partes do centro e norte da ilha principal. A metade sul é menos habitada. Muitos pescadores comerciais pequenos povoam a ilha. Os barcos de pesca, as rendeiras, o folclore, a culinária e a arquitetura colonial contribuem para o

crescimento do turismo e atraem recursos que compensam a falta de um grande parque industrial. Vilarejos imersos em tradição e história, como Santo Antônio de Lisboa e Ribeirão da Ilha, ainda resistem aos avanços da modernidade.

Figura 10: Figura Representativa de Florianópolis - Santa Catarina



Fonte: <[litoraldesantacatarina.com.br](http://litoraldesantacatarina.com.br)>. Acesso em 05 nov. 2019

### 5.3.2 Breve Histórico de Florianópolis

Segundo a Secretaria de Turismo (2019), os primeiros habitantes da região de Florianópolis foram os índios tupis-guaranis. Praticavam a agricultura, mas tinham na pesca e coleta de moluscos as atividades básicas para sua subsistência.

Os indícios de sua presença encontram-se nos sambaquis e sítios arqueológicos cujos registros mais antigos datam de 4.800 a.C.

Já no início do século XVI, embarcações que demandavam à Bacia do Prata aportavam na Ilha de Santa Catarina para abastecerem-se de água e víveres. Entretanto, somente por volta de 1675 é que Francisco Dias Velho, junto com sua família e agregados, dá início a povoação da ilha com a fundação de Nossa Senhora do Desterro - segundo núcleo de povoamento mais antigo do Estado, ainda fazendo parte da vila de Laguna - desempenhando importante papel político na colonização.

A partir desta data intensifica-se o fluxo de paulistas e vicentistas que ocupam vários outros pontos do litoral. Em 1726, Nossa Senhora do Desterro é elevada a categoria de vila, a partir de seu desmembramento de Laguna.

Com a ocupação, tiveram prosperidade a agricultura e a indústria manufatureira de algodão e linho, permanecendo resquícios desse passado no que se refere à confecção artesanal da farinha de mandioca e das rendas de bilro.

Nesta época, em meados do século XVIII, verifica-se a implantação das "armações" para pesca da baleia, em Armação da Piedade (Governador Celso Ramos) e Armação do Pântano do Sul (Florianópolis), cujo óleo era comercializado pela Coroa fora do estado de Santa Catarina.

No século XIX, Desterro foi elevada à categoria de cidade; tornou-se Capital da Província de Santa Catarina em 1823 e inaugurou um período de prosperidade, com o investimento de recursos federais. Projetou-se a melhoria do porto e a construção de edifícios públicos, entre outras obras urbanas. A modernização política e a organização de atividades culturais também se destacaram, marcando inclusive os preparativos para a recepção ao Imperador D. Pedro II (1845).

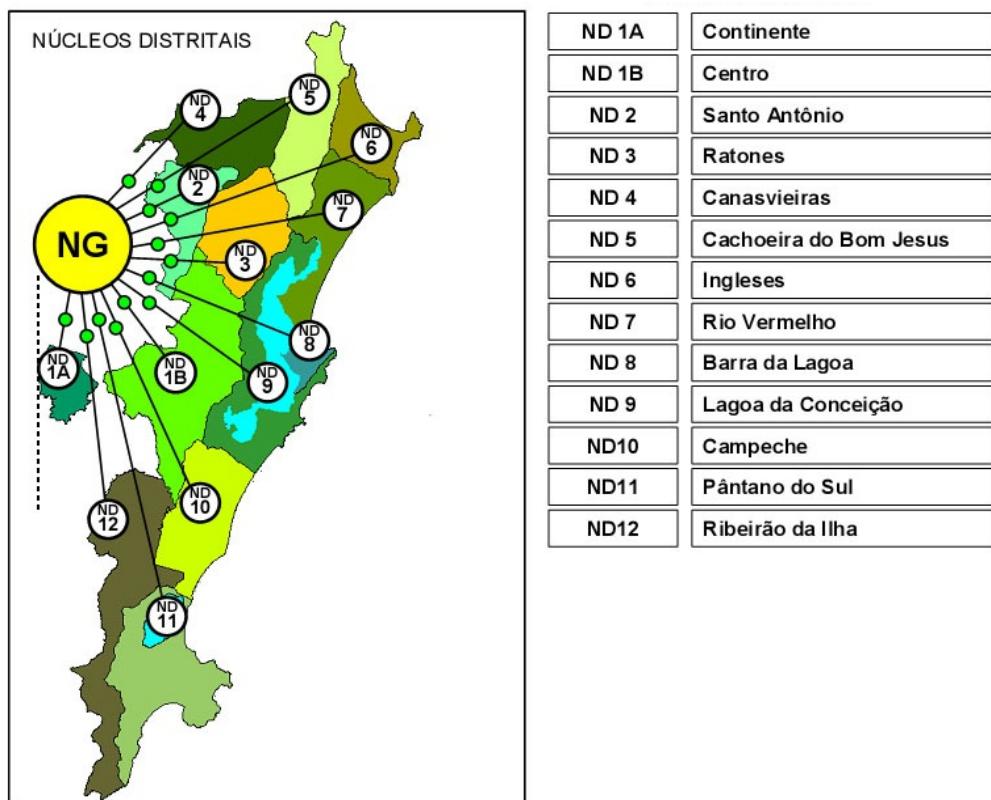
Com o advento da República (1889), as resistências locais ao novo governo provocaram um distanciamento do governo central e a diminuição dos seus investimentos. A vitória das forças comandadas pelo Marechal Floriano Peixoto determinaram em 1894 a mudança do nome da cidade para Florianópolis.

### **5.3.3 Geografia de Florianópolis**

A geografia de Florianópolis, compreendida a área da parte continental e da ilha, abrange 675,410 km<sup>2</sup>, com uma população de 500.973 habitantes em 2019, segundo estimativas do IBGE. Fazem parte do município de Florianópolis os

seguintes distritos: Sede (Centro e Continente), Santo Antônio de Lisboa, Ratones, Canasvieiras, Cachoeira do Bom Jesus, Ingleses, Rio Vermelho, Barra da Lagoa, Lagoa da Conceição, Campeche, Pântano do Sul e Ribeirão da Ilha.

Figura 11: Localização dos Núcleos Distritais de Florianópolis



Fonte: <urbanidades.arc.br>. Acesso em 05 nov. 2019

Florianópolis é uma das três ilhas-capitais do Brasil, localiza-se no leste do estado de Santa Catarina e é banhada pelo oceano Atlântico. Grande parte de seu território, aproximadamente 97,23%, está situado na ilha de Santa Catarina.

Varrida por ventos muito variáveis, possui um clima subtropical úmido, que se caracteriza pela alternância de verões e invernos, e farta distribuição anual de chuvas. Isto em conjunto com suas 42 praias, contribuiu para que ela se tornasse a capital turística do Mercosul, pois possui um intenso movimento turístico durante todo o verão, principalmente com argentinos, gaúchos e paulistas.

Existem atualmente doze distritos na região do município de Florianópolis, entre eles o centro. O distrito central ou distrito sede de Florianópolis engloba a parte da Ilha de Santa Catarina situada mais próxima ao estreito que a separa do continente, bem como a própria região continental do município.

É a região mais populosa da cidade e a que concentra o maior número de pontos turísticos não-naturais. É no distrito central de Florianópolis que se encontram os bairros objeto dessa pesquisa: Centro, Agronômica e Trindade.

**Bairro Centro:** O Centro é um bairro nobre da região central de Florianópolis. Apesar de estar no oeste da Ilha de Santa Catarina, o Centro está geograficamente no centro da área urbana da Grande Florianópolis e é área mais densa e verticalizada da Ilha de Santa Catarina. Foi do Centro que partiu a povoação da cidade, e é ali que ficam, por exemplo, a Prefeitura da cidade, a Câmara de Vereadores, a Assembleia Legislativa, entre outros. A avenida mais conhecida do Centro é a Avenida Beira-Mar Norte que, além de ciclovia e calçadão, possui alguns dos apartamentos mais caros da cidade. Essa avenida é complementada, ao sul, pela avenida Paulo Fontes, e pela via entre as pontes e o Túnel Antonieta de Barros.

**Bairro Agronômica:** A Agronômica é um bairro da região central de Florianópolis. Está situado na porção do município situada na Ilha de Santa Catarina, ao norte do Morro da Cruz, entre os bairros da Trindade e o Centro. Na Agronômica está situado o Palácio da Agronômica, residência oficial do governador do Estado. No bairro também se encontra o Hospital Infantil Joana de Gusmão e o Hospital Nereu Ramos, a 6<sup>a</sup> DP (Delegacia de Proteção à Mulher, ao Menor e ao Adolescente) e também a Escola de Educação Básica Padre Anchietta.

**Bairro Trindade:** A Trindade é um bairro na região central de Florianópolis. Localizado entre os pés do maciço do Morro da Cruz e o Manguezal do Itacorubi, é o segundo mais populoso da cidade, fazendo divisa com os bairros Córrego Grande, Agronômica, Santa Mônica, Pantanal e Itacorubi. É conhecido como um bairro universitário, por sediar o principal campus da Universidade Federal de Santa Catarina (UFSC). Originário de uma freguesia ocupada por imigrantes açorianos desde o século XVIII, desenvolveu-se nas últimas décadas em torno da Universidade Federal de Santa Catarina. Devido à sua localização em relação ao centro da antiga Desterro era denominado na época imperial como Trás-do-Morro. Além do campus da UFSC, no bairro ficam a Academia de Polícia Militar, um batalhão do Corpo de Bombeiros e um terminal de ônibus (TITRI), que atende a região central de Florianópolis, na ilha de Santa Catarina.

Figura 12: Bairros Estudados em Florianópolis



Fonte: <googlemaps.com.br>. Acesso em 05 nov. 2019

#### **5.3.4 Aspectos Sócio-Econômicos**

Segundo a Secretaria de Turismo (2019), a economia de Florianópolis é voltada para os setores público, comércio e de serviços, com foco principal no turismo. A indústria não é forte por questões ambientais, fazendo com que este setor se desenvolva nas cidades da macrorregião de Florianópolis. Apesar disso, a capital é destaque quando se fala em tecnologia. Quase 300 empresas de ponta formam o parque tecnológico da cidade, fornecendo tecnologia até para o mercado internacional. O faturamento deste setor chega a R\$ 500 milhões ao ano.

Conforme o sítio Governo de Santa Catarina (2019), a Grande Florianópolis destaca-se nos setores de tecnologia, turismo, serviços e construção civil. O Norte do estado é polo tecnológico, moveleiro e metal-mecânico. O Oeste concentra atividades de produção alimentar e de móveis. O Planalto Serrano tem a indústria de papel, celulose e da madeira. O Sul destaca-se pelos segmentos do vestuário, plásticos descartáveis, carbonífero e cerâmico. No Vale do Itajaí, predomina a indústria têxtil, naval e de tecnologia. O turismo é forte na economia catarinense.

Para atender às demandas de seu público acadêmico, a cidade investiu pesadamente em estradas e escolas, sendo agora bem classificada em diferentes aspectos do desenvolvimento, como alfabetização (97%) e eletrificação (~100%).

## 5.4 COLETA DE DADOS DO MERCADO IMOBILIÁRIO

Conforme Gerhardt e Silveira (2009), a coleta de dados é a busca por informações para a elucidação do fenômeno ou fato que o pesquisador quer desvendar. O instrumental técnico elaborado pelo pesquisador para o registro e a medição dos dados poderá ser de diversas formas, estando entre elas a:

**Pesquisa Eletrônica:** é constituída por informações extraídas de endereços eletrônicos, disponibilizados em *home pages* e sítios, a partir de livros, folhetos, manuais, guias, artigos de revistas, artigos de jornais, etc. Apesar de sua importante contribuição para a investigação científica, nem toda informação disponibilizada em meios eletrônicos deve ser considerada como sendo de caráter científico.

Para esse trabalho fez-se o levantamento de dados de mercado junto às imobiliárias que ofertam imóveis na região de estudo, utilizando-se para isso, os anúncios apresentados em suas páginas na internet e, quando necessário, entrava-se em contato com o corretor de imóveis para sanear as dúvidas mais pontuais.

Ressalta-se que, segundo a NBR 14.653-2 (2011), o número de dados de mercado que compõe a amostra, quando se utiliza o método comparativo direto de dados de mercado, é função do número de variáveis explicativas efetivamente utilizadas no modelo de regressão e do grau de fundamentação que se pretende atingir. Para se atingir o Grau III de fundamentação, por exemplo, um dos requisitos que a norma exige é que se tenha pelo menos  $6 \cdot (k + 1)$  dados de mercado na amostra, onde  $k$  é o número de variáveis independentes do modelo de regressão.

A coleta dos dados utilizados como amostra para essa pesquisa foi realizada entre os meses de março e abril de 2020, nos bairros Centro, Agronômica e Trindade, em Florianópolis/SC, e continha, inicialmente, uma base com 225 dados de oferta de apartamentos desses três bairros. Nesse caso, especificamente para a Engenharia de Avaliações, o número de dados coletados permitiria, respeitando a NBR 14.653-2, efetuar uma modelagem com aproximadamente 30 variáveis explicativas e ainda assim atingir o Grau III de fundamentação para esse item da norma.

Desta base, retirou-se uma amostra aleatória contendo dados de 35 imóveis, para compor a amostra de validação. Para definição dos dados que fariam parte da amostra de validação utilizou-se a função =ALEATÓRIOENTRE(0;225) no excel.

Os dados de mercado coletados para o estudo (amostra de trabalho) com os respectivos polos de valorização, podem ser vistos espacialmente na Figura 13.

Figura 13: Distribuição espacial dos dados de mercado da amostra de trabalho



Fonte: o autor

Destaca-se que os pontos marcados em vermelho na figura referem-se a supostos polos de valorização dos imóveis, a saber: 1) ponto que representa o possível polo de valorização Praça XV de Novembro; 2) ponto que representa o possível polo de valorização Shopping Beira Mar; 3) ponto que representa o possível polo de valorização Shopping Iguatemi; 4) ponto que representa o possível polo de valorização Centro Tecnológico (CTC) na UFSC; 5) linha que representa o possível polo de valorização referente à Avenida Beira Mar Norte, de Florianópolis. Salienta-se que na área em estudo não há nenhum polo de desvalorização significativo e, por isso, não foram incorporadas variáveis com essa característica nas modelagens.

A Tabela 4 mostra a distribuição das amostras coletadas em cada bairro, chamada aqui de amostra de trabalho, com a respectiva amostra de validação.

Tabela 4: Quantidade de amostra de trabalho e de validação em cada bairro

Bairro	Amostra de Trabalho	Amostra de Validação
Centro	90	17
Agronômica	45	6
Trindade	55	12
<b>Total</b>	<b>190</b>	<b>35</b>

Fonte: o autor

Os dados de mercado da amostra de validação, selecionados aleatoriamente no excel, podem ser vistos, distribuídos espacialmente, na Figura 14.

Figura 14: Distribuição espacial dos dados da amostra de validação



Fonte: o autor

Uberti (2016) utilizou uma amostra de validação contendo 18% dos dados totais da amostra, Malaman (2014) utilizou como amostra de validação 8% dos dados totais coletados e Faria Filho (2015) utilizou uma amostra de validação correspondente a 10% da amostra total de dados. Cesare (2012) sugere extrair da amostra total um percentual entre 15% e 25% para compor a amostra de validação de uma avaliação em massa. Optou-se, nesse estudo, pela utilização de 35 dados de validação, que corresponde a 16% dos dados totais da amostra.

#### 5.4.1 Descrição das Variáveis da Amostra

Inicialmente, como limitação para este trabalho, destaca-se o fato de que algumas variáveis comumente utilizadas em engenharia de avaliações, como a idade, o estado de conservação e o andar em que determinado imóvel se encontra, não foram consideradas nesse estudo. Em função da base de dados ser relativamente grande, a visita aos imóveis, um a um, tornou-se inviável. Soma-se a isso a necessidade de isolamento social para enfrentamento da pandemia de Covid-19. Contudo, dentro de certa aleatoriedade, espera-se que o modelo capte o valor médio dessas variáveis. Os dados coletados foram agrupados conforme a tabela do Apêndice 01, e as características coletadas para cada imóvel são descritas abaixo.

Para o comportamento das variáveis dependentes e independentes dos dados da amostra, em condições *ceteris paribus*, é esperado conforme descrito:

### **Variáveis Dependentes:**

Valor Total (VT): Variável dependente quantitativa contínua, expressa em unidades monetárias (R\$), com o objetivo de indicar qual é o valor total do imóvel, em reais. Para representar a variável dependente utilizou-se o valor unitário (VU), pela possibilidade deste oferecer mais detalhes para se extrair informações e conclusões;

Valor Unitário (VU): Variável dependente quantitativa contínua, expressa em unidades monetárias por metro quadrado de área (R\$/m<sup>2</sup>), com o objetivo de indicar qual é o valor, nesse caso em reais, de cada unidade de área do imóvel em questão;

### **Variáveis Independentes:**

Área Privativa (AP): É uma variável quantitativa contínua que representa a área do imóvel, excluindo-se a sua área comum (espaço de propriedade de todos), expressa em metros quadrados (m<sup>2</sup>). Como comportamento da variável espera-se que quanto maior a área privativa de um imóvel, menor será o seu valor unitário;

Número de Dormitórios (ND): Diz respeito à quantidade total de dormitórios existentes em dado imóvel, juntamente com as suítes. Essa variável é quantitativa discreta e espera-se que, com o aumento dos dormitórios, o valor unitário aumente;

Número de Banheiros (NB): Variável quantitativa discreta e diz respeito à quantidade total de banheiros existentes em dado imóvel. É uma variável quantitativa. Com o aumento do total de banheiros, o valor unitário de um imóvel tende a aumentar;

Número de Vagas de Garagem (NG): Número de vagas para estacionamento dos veículos de cada imóvel, coberta e descoberta, com ou sem matrícula. Trata-se de uma variável quantitativa discreta. Como comportamento espera-se um aumento do valor do imóvel, quando houver um aumento do número total de garagens;

Número de Suítes (NS): Essa variável refere-se ao número de quartos que contenham um banheiro privativo. É uma variável quantitativa discreta e espera-se que, aumentando-se o total de suítes de um imóvel, ocorra a valorização do mesmo;

Distância ao Polo Praça XV (DPXV): Refere-se à distância real, em metros (m), entre o imóvel coletado e um ponto central localizado na praça XV de novembro, em Florianópolis. Essa variável representa a rota mais rápida para ir de um ponto ao outro, de automóvel. Tem-se aqui uma variável quantitativa contínua. O comportamento que se espera para esta variável é que o aumento da distância do imóvel em relação ao polo Praça XV cause uma redução do valor unitário do imóvel;

Distância ao Polo Shopping Beira Mar (DSBM): Refere-se à distância real, em metros (m), entre o imóvel coletado e o Shopping Beira Mar, em Florianópolis. Essa variável representa a rota mais rápida para ir de um ponto ao outro, de automóvel. Tem-se aqui uma variável quantitativa contínua. O comportamento que se espera para esta variável é que o aumento da distância do imóvel coletado em relação polo Shopping Beira Mar provoque uma redução do valor unitário do imóvel;

Distância ao Polo Shopping Iguatemi (DSIG): Refere-se à distância real, em metros (m), entre o imóvel coletado e o Shopping Iguatemi, em Florianópolis. Essa variável representa a rota mais rápida para ir de um ponto ao outro, de automóvel. Tem-se aqui uma variável quantitativa contínua. O comportamento que se espera para esta variável é que o aumento da distância do imóvel coletado em relação polo de referência (Shopping Iguatemi) provoque uma redução do valor unitário do imóvel;

Distância ao Polo CTC UFSC (DCTC): Refere-se à distância real, em metros (m), entre o imóvel coletado e um ponto de referência (fachada central) do Centro Tecnológico (CTC) da UFSC. Essa variável representa a rota mais rápida para ir de um ponto ao outro, de automóvel. Tem-se aqui uma variável quantitativa contínua. O comportamento esperado é que o aumento da distância do imóvel em relação ao Centro Tecnológico (CTC) provoque uma redução do valor unitário do imóvel;

Distância ao Polo Beira Mar (DABM): Refere-se à distância real, em metros (m), entre o imóvel coletado e a faixa da Avenida Beira Mar de Florianópolis/SC. Tem-se

aqui uma variável quantitativa contínua. O comportamento que se espera para esta variável é que o aumento da distância do imóvel em relação ao polo de referência (Beira Mar) provoque uma redução do valor unitário do imóvel;

Mobiliado (MO): Variável que representa se o apartamento é sem mobília, é semi-mobiliado ou mobiliado. Para essa variável utilizou-se códigos alocados com as seguintes codificações: sem mobília = valor 1; semi-mobiliado = valor 2; mobiliado = valor 3. Espera-se um aumento dos valores quando existe mobília;

Piscina (PS): Variável dicotômica que busca identificar a variabilidade dos valores dos imóveis em razão da existência de piscina, sendo 0 (zero) se não tiver e 1 (um) se tiver. Espera-se que a existência de piscina aumente o valor unitário do imóvel;

Churrasqueira (CH): Variável dicotômica que busca identificar a variabilidade dos valores dos imóveis em razão da existência de churrasqueira no apartamento, sendo que a codificação é 0 (zero) se não há e 1 (um) caso haja churrasqueira. Espera-se que existência de sacada com churrasqueira agregue valor ao imóvel em questão;

Padrão Construtivo (PC): O padrão de construção das edificações coletadas representa de maneira qualitativa os três padrões construtivos levados em consideração nesse estudo. Para essa variável utilizou-se códigos alocados com as seguintes codificações: padrão baixo = valor 1; padrão médio = valor 2; padrão alto = valor 3. Espera-se, para esta variável, um aumento do valor unitário do imóvel à medida que ocorre o aumento do padrão construtivo;

Coordenada UTM E (E): Representa a coordenada métrica (Universal Transversa de Mercator – UTM) da amostra representando a localização geográfica E do respectivo imóvel, sendo que o *datum* usado nesse estudo foi o WGS84;

Coordenada UTM N (N): Representa a coordenada métrica (Universal Transversa de Mercator – UTM) da amostra representando a localização geográfica N do respectivo imóvel, sendo que o *datum* usado nesse estudo foi o WGS84;

Nome do Edifício e Endereço: Informação sobre o nome do edifício e seu endereço, com o nome da rua e o número, representando a localização física do imóvel.

## 6 RESULTADOS E ANÁLISES

Inicialmente é realizada a análise exploratória dos dados da amostra e na sequência são apresentados os modelos de regressão linear múltipla, de regressão espacial e de regressão geograficamente ponderada obtidos para os dados de mercado coletados. É apresentado, também, os cálculos relativos às medidas de qualidade do ajuste e desempenho dos modelos. Finalmente, são apresentadas as superfícies de valores unitários interpolados para a área em estudo.

### 6.1 ANÁLISE EXPLORATÓRIA DOS DADOS DA AMOSTRA

Objetivando avaliar a existência de possíveis valores atípicos na distribuição e obter maior familiaridade com os dados da amostra, será realizada uma análise exploratória dos dados, corrigindo eventuais problemas para que, na fase de modelagem, se obtenha o modelo de regressão mais ajustado aos dados coletados.

Realizou-se a análise exploratória dos dados tanto na variável explicada quanto nas variáveis explicativas, utilizando como ferramentas estatísticas os gráficos de *boxplot* e de dispersão, momentos, correlações parciais e histogramas.

O roteiro de pesquisa que será seguido neste tópico, com adaptações ao objetivo deste trabalho, é aquele sugerido no trabalho de Uberti (2016).

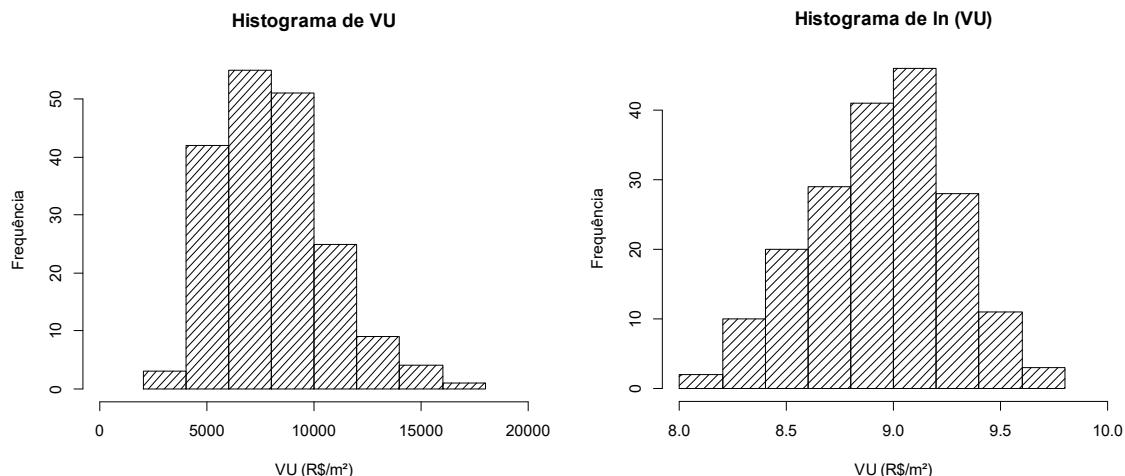
#### 6.1.1 Análise da Variável Explicada (Valor Unitário)

Esta análise preliminar da variável valor unitário tem como objetivo verificar se há a necessidade de alguma espécie de transformação em sua escala. Realizou-se a comparação entre a escala original (VU) e a transformação logarítmica natural ( $\ln VU$ ). Segundo Dantas (2012), a transformação logarítmica é a preferida quando se procura ajustar modelos a dados de valores imobiliários.

Para Uberti (2016), é coerente a utilização de transformação logarítmica uma vez que as variáveis explicadas possuindo valores no campo dos reais positivos garantem que o campo de variação dos valores ajustados correspondentes também serão os reais positivos. A autora cita ainda que a transformação logarítmica na variável explicada torna o modelo multiplicativo, bem como pode estabilizar a variância do modelo, característica sugerida pelas normas brasileiras de avaliações.

Construiu-se dois histogramas para a variável independente (valor unitário), um na escala original (VU) e outro na escala logaritmo natural ( $\ln VU$ ) com o objetivo de analisar o comportamento dessa variável conforme a Figura 15.

Figura 15: Histograma de frequências de VU e de  $\ln (VU)$



Fonte: o autor

Pode-se verificar que na escala original os dados apresentavam leve padrão de assimetria positiva, com os dados assimétricos para direita. Ao se realizar o histograma da variável transformada em logaritmo natural, pode-se verificar uma correção dessa assimetria, tornando os dados mais próximos da curva normal.

Para os dados apresentarem assimetria positiva, deve-se ter a relação  $Mo < Me < \bar{x}$ . Na Tabela 5 abaixo pode-se verificar os valores desses parâmetros.

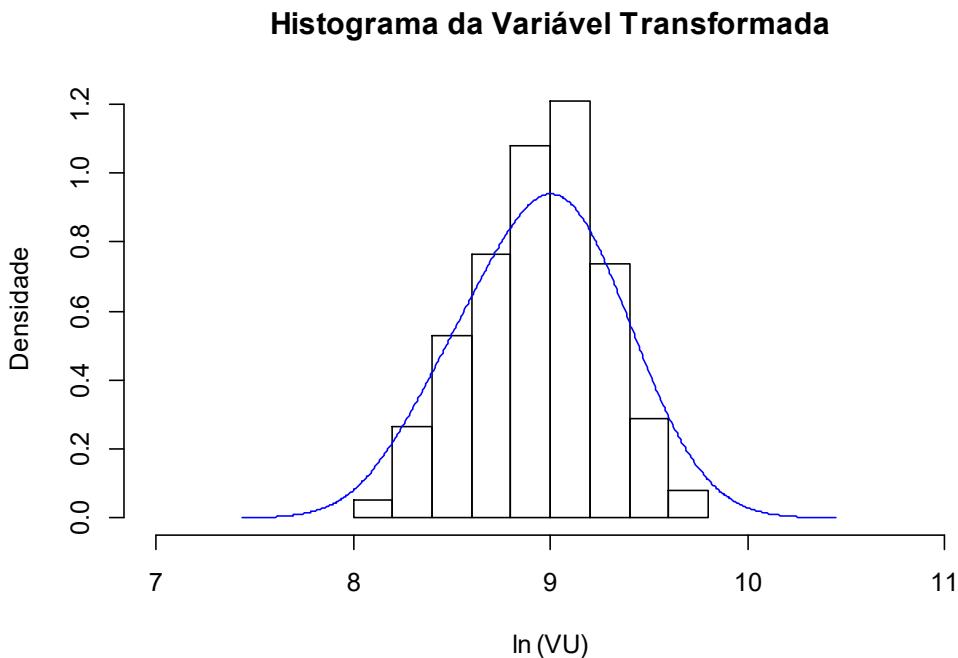
Tabela 5: Medidas de tendência central da variável explicada

	<b>Moda (Mo)</b>	<b>Mediana (Me)</b>	<b>Média (<math>\bar{x}</math>)</b>
<b>VU</b>	7.840,00	7.848,50	8.105,25
<b>In (VU)</b>	8,967	8,968	8,947

Fonte: o autor

Segundo Dantas (2012), os dados imobiliários, em geral, tendem a se concentrar em torno de um valor central. Dependendo da dispersão dos dados observados, essa concentração pode ser maior ou pode ser menor.

Figura 16: Histograma da variável transformada  $\ln(VU)$



Fonte: o autor

Para confirmar essa situação, calculou-se os coeficientes de assimetria e curtose para os dados da amostra nas duas situações: sem transformação e com transformação logaritmo natural. Os resultados constam na Tabela 6:

Tabela 6: Coeficientes de assimetria e curtose da variável explicada

	Assimetria ( $A$ )	Curtose ( $C$ )
VU	0,56	3,02
$\ln(VU)$	-0,17	2,37

Fonte: o autor

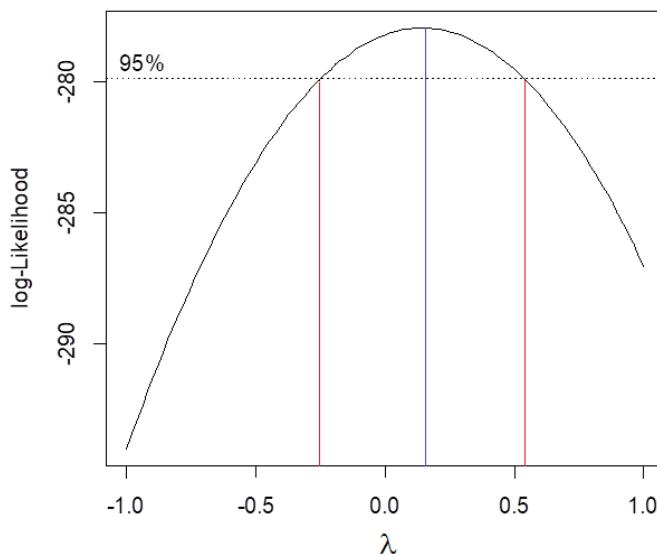
Com relação à simetria, tem-se assimetria positiva se  $A > 0$ , assimetria negativa se  $A < 0$  e simetria se  $A = 0$ . Segundo Gujarati *et al.* (2011, p. 811), com relação à curtose, tem-se curvas leptocúrticas quando os dados são bem concentrados ( $C > 3,26$ ) e a distribuição é mais afunilada com pico mais alto que a distribuição normal; curvas platicúrticas quando os dados são espalhados ( $C < 3,26$ ) e a distribuição é mais achatada que a distribuição normal e mesocúrticas no caso intermediário ( $C = 3,26$ ), em que a distribuição tem o achatamento da distribuição normal.

Quando a variável explicada valor unitário (VU) é mantida na escala original, tem-se uma distribuição assimétrica positiva e platicúrtica. Já quando a variável valor unitário passa por transformação logaritmo natural ( $\ln VU$ ), tem-se uma curva de assimetria levemente negativa, verificado pela Figura 16, e platicúrtica, com a função de distribuição mais achatada e pico mais baixo que a distribuição normal.

Com o objetivo de confirmar se essa transformação se mostraria adequada ao conjunto de dados coletados, realizou-se uma análise gráfica Box & Cox.

Segundo Box e Cox (1964), o valor do parâmetro lambda definirá qual a transformação mais adequada para a variável em análise, sendo que, para valores positivos, tem-se  $\ln y$  se  $\lambda = 0$ ,  $y^{0.5}$  se  $\lambda = 0,5$ ,  $y$  se  $\lambda = 1,0$  e  $y^2$  se  $\lambda = 2,0$ .

Figura 17: Gráfico Box-Cox do valor unitário não transformado



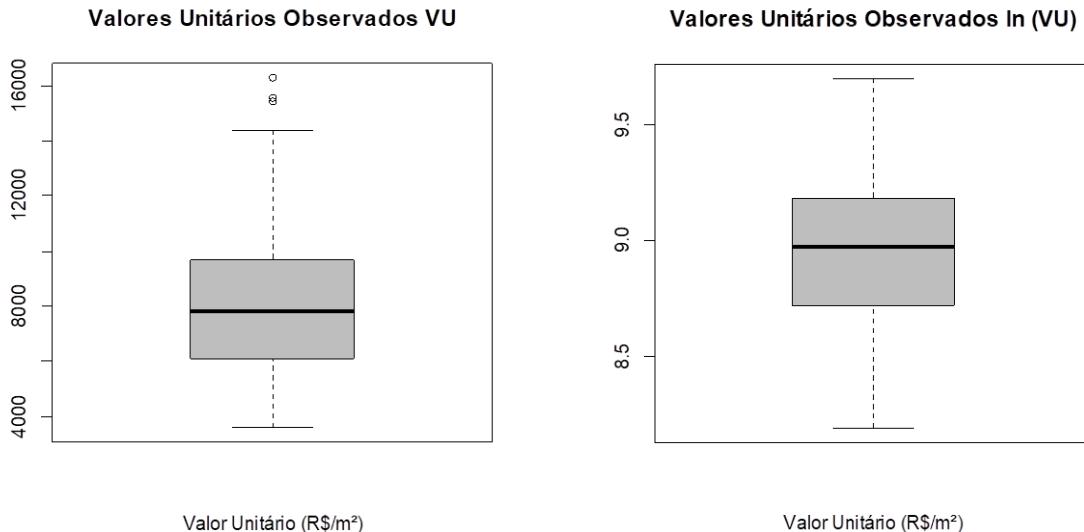
Fonte: o autor

Os dados desse estudo apresentaram o valor  $\lambda = 0,15$ , indicando que dentre as transformações possíveis, o logaritmo natural fornecerá resultados bem próximos dos obtidos com a transformação Box & Cox (1964) dada por  $y' = (y^\lambda - 1)/\lambda$  se  $\lambda \neq 0$ .

O gráfico da Figura 17 mostra ainda um intervalo de confiança de 95% para o valor de  $\lambda$  e percebe-se que o 0 está contido neste intervalo. Isto novamente confirma que a transformação  $\ln (VU)$  é a mais adequada para o conjunto de dados.

Construiu-se também, para a variável explicada, na escala original e transformada em logaritmo natural, gráficos *boxplot* com o objetivo de verificar a existência de pontos discrepantes que estejam influenciando os resultados.

Figura 18: Gráfico *boxplot* com a variável VU e  $\ln(VU)$



Fonte: o autor

Conforme a Figura 18, o gráfico *boxplot* para a variável valor unitário (VU) em escala original apresentou três *outliers*, pontos esses considerados discrepantes e que acabam por afetar a distribuição na parte superior dos quartis. Os três pontos considerados *outliers* são aqueles cujos valores unitários são R\$ 15.559,00, R\$ 15.426,00 e R\$ 16.290,00. Observa-se que, quando a distribuição passava por transformação logaritmo natural, os dados que antes eram considerados *outliers* deixam de ser e tem-se, então um conjunto de dados ( $\ln VU$ ) melhor distribuídos.

Realizou-se também a análise das estatísticas descritivas para a variável explicada na forma original (VU) e com transformação logaritmo natural ( $\ln VU$ ). Os resultados dessas estatísticas podem ser visualizados na Tabela 7:

Tabela 7: Estatísticas descritivas da variável explicada VU e  $\ln (VU)$

Estatística	VU (R\$/m <sup>2</sup> )	$\ln (VU)$ (R\$/m <sup>2</sup> )
Média	8.105,25	8,947
1º Quartil	6.107,00	8,717
Mediana	7.848,50	8,968
3º Quartil	9.682,00	9,178
Moda	7.840,00	8,967
Desvio Padrão	2.645,38	0,332
Variância	6.998.049,67	0,111
Curtose	3,02	2,37

Assimetria	0,56	-0,169
Amplitude	12.682,00	1,507
Mínimo	3.608,00	8,191
Máximo	16.290,00	9,698
Contagem	190	190
CV (%)	32,63%	3,70%

Fonte: o autor

Percebe-se, para a variável sem transformação (VU), um coeficiente de variação na ordem de 32,63%. Esse valor mais acentuado deve-se à grande amplitude dos valores unitários dos imóveis que possuem um mínimo de R\$ 3.608,00/m<sup>2</sup> e um máximo de R\$ 16.290,00/m<sup>2</sup>, o que não ocorre quando é realizada a transformação logaritmo natural, apresentando um coeficiente de variação baixo.

Por todas essas características e por não apresentar, de início, *outliers* em sua distribuição, adotou-se a transformação logaritmo natural na variável valor unitário (ln VU) para modelagem dos dados por regressão linear múltipla.

### 6.1.2 Análise das Variáveis Explicativas

A NBR 14.653-2 (2011) diz que uma forte dependência linear entre duas ou mais variáveis independentes provoca degenerações no modelo e limita a sua utilização. As variâncias das estimativas dos parâmetros podem ser muito grandes e acarretar a aceitação da hipótese nula e a eliminação de variáveis fundamentais.

Segundo a norma, para verificação da multicolinearidade deve-se, em primeiro lugar, analisar a matriz das correlações, que espelha as dependências lineares de primeira ordem entre as variáveis independentes, com atenção especial para resultados superiores a 0,80 e, em caso de correlação linear elevada entre quaisquer subconjuntos de variáveis independentes, isto é, multicolinearidade, deve-se examinar a coerência das características do imóvel avaliando com a estrutura de multicolinearidade inferida, vedada a utilização do modelo em caso de incoerência.

Para se fazer essa análise nas variáveis explicativas, utilizou-se a matriz de correlações com o objetivo de detectar possíveis variáveis explicativas correlacionadas e como se dá a correlação delas com a variável explicada.

A matriz de correlações a seguir foi calculada com as variáveis em escala original. Finalizada essa análise nas variáveis explicativas, procedeu-se o estudo de

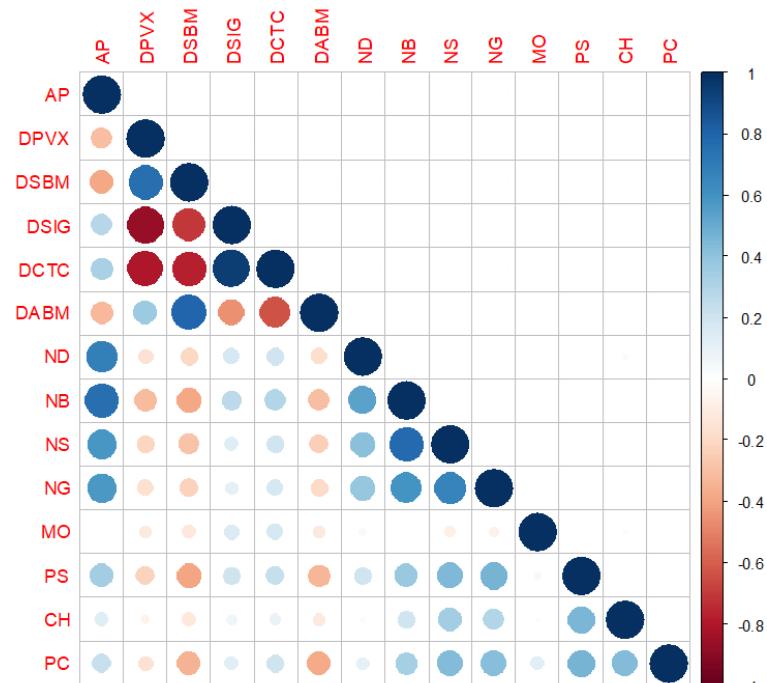
possíveis transformações em suas escalas, buscando manter a coerência dos regressores e objetivando melhorar o modelo de regressão linear múltipla, para que ele possa explicar com fidelidade o mercado imobiliário em estudo.

Tabela 8: Matriz de correlações das variáveis explicativas

	AP	DPVX	DSBM	DSIG	DCTC	DABM	ND	NB	NS	NG	MO	PS	CH	PC
<b>AP</b>	1,00													
<b>DPVX</b>	-0,31	1,00												
<b>DSBM</b>	-0,38	0,76	1,00											
<b>DSIG</b>	0,29	<b>-0,87</b>	-0,71	1,00										
<b>DCTC</b>	0,32	<b>-0,81</b>	-0,78	<b>0,94</b>	1,00									
<b>DABM</b>	-0,32	0,37	0,79	-0,45	-0,63	1,00								
<b>ND</b>	0,65	-0,20	-0,30	0,20	0,25	-0,24	1,00							
<b>NB</b>	0,76	-0,31	-0,38	0,26	0,30	-0,30	0,56	1,00						
<b>NS</b>	0,59	-0,21	-0,29	0,14	0,19	-0,24	0,48	0,77	1,00					
<b>NG</b>	0,57	-0,17	-0,22	0,12	0,18	-0,19	0,53	0,59	0,66	1,00				
<b>MO</b>	0,01	-0,12	-0,13	0,16	0,17	-0,12	0,08	0,00	-0,08	-0,07	1,00			
<b>PS</b>	0,35	-0,22	-0,40	0,20	0,24	-0,33	0,35	0,37	0,45	0,46	-0,05	1,00		
<b>CH</b>	0,14	-0,06	-0,13	0,07	0,09	-0,11	0,14	0,19	0,35	0,30	0,01	0,46	1,00	
<b>PC</b>	0,22	-0,16	-0,35	0,12	0,21	-0,37	0,35	0,34	0,44	0,43	0,13	0,47	0,43	1,00

Fonte: o autor

Figura 19: Diagrama de cores indicando as correlações entre as variáveis



Fonte: o autor

Na Figura 19, quanto maior o tamanho do círculo e mais escura a cor, maior o nível de correlação entre as variáveis correspondentes, sendo azul uma correlação do tipo positiva e vermelho uma correlação do tipo negativa.

Pode-se perceber que a variável distância ao shopping Iguatemi (DSIG) e distância ao Centro Tecnológico (DCTC) estão fortemente correlacionadas, muito provavelmente por serem pontos próximos e representarem a mesma região.

As variáveis distância ao shopping Beira Mar (DSBM) e distância à Avenida Beira Mar (DABM) apresentaram correlação na ordem de 0,80, que é o limite sugerido pela norma. É provável que essa correlação se deva, em partes, a mesma influência que ambos polos têm sobre o valor dos imóveis do entorno.

Realizou-se, então, simulações para verificar quais dessas variáveis tinham maior influência e como se daria a nova matriz de correlações após a exclusão de uma ou mais variáveis dessas correlacionadas. Nas diversas combinações possíveis, o modelo clássico tornou-se incoerente e acabou por não captar, de forma simultânea, a influência de dois ou mais supostos polos de valorização.

É provável que isso tenha ocorrido em virtude de, na amostra, os imóveis que apresentam maiores valores unitários se encontram no bairro Centro e, nesse caso, estarem distantes do bairro Trindade, onde se encontra a UFSC, por exemplo.

Os cinco supostos polos de valorização (DPXV, DSBM, DSIG, DCTC, DABM) foram coletados com o objetivo de verificar quais tinham maior influência sobre o valor dos imóveis na área em estudo e em uma análise preliminar, parte deles não se mostraram significantes ao nível de 10%, como visto na Tabela 9:

Tabela 9: Significâncias das variáveis excluídas da fase de modelagem

Variável	$t_{\text{calc}}$	Significância
Distância à Praça XV (DPXV)	0,276	0,390
Distância ao Shopping Beira Mar (DSBM)	-0,346	0,360
Distância ao Shopping Iguatemi (DSIG)	0,437	0,330
Distância ao Centro Tecnológico (DCTC)	0,691	0,250
Distância à Avenida Beira Mar (DABM)	-0,766	0,220

Fonte: o autor

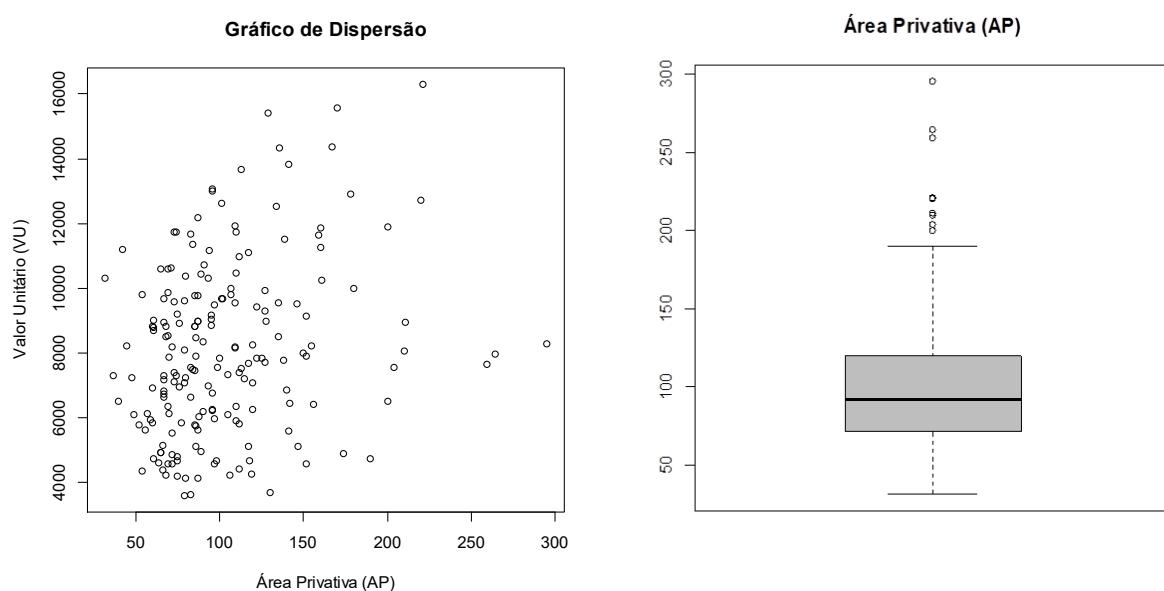
Optou-se, então, pela exclusão das variáveis distância ao shopping Iguatemi (DSIG), distância ao shopping Beira Mar (DSBM), distância ao Centro Tecnológico

(DCTC) e distância à Praça XV (DPXV) da fase de modelagem da regressão clássica, permanecendo, desta forma, a variável distância à Avenida Beira Mar (DABM), por ter se mostrado a mais influente e significativa dentre as variáveis de localização. Convém ressaltar que a variável distância à Avenida Beira Mar (DABM) tornava-se significante ao nível de 10% quando deixada no modelo de regressão.

Das variáveis explicativas restantes, procedeu-se a construção dos gráficos de dispersão da variável valor unitário (VU), versus variáveis independentes, todos em escala original e de gráficos *boxplot* com os respectivos coeficientes de correlação. Essa análise permitirá compreender como se dá a relação entre a variável dependente e as variáveis independentes, possibilitando decidir se a variável seguirá para modelagem ou se precisará de alguma transformação.

Inicialmente construiu-se o gráfico de dispersão das variáveis área privativa (AP) e valor unitário (VU), em escala original. Construiu-se, também, o gráfico *boxplot* dos dados de área privativa e essa ferramenta permitiu detectar pontos atípicos no quartil superior da distribuição. Esses gráficos possibilitaram perceber que uma transformação na variável área privativa pode ser necessária. O coeficiente de correlação entre a variável área privativa (AP) e valor unitários (VU) é  $r = 0,27$ . Essa variável permanecerá nos cálculos por ser importante na fase de modelagem.

Figura 20: Gráfico de dispersão e *boxplot* da variável área privativa (AP)

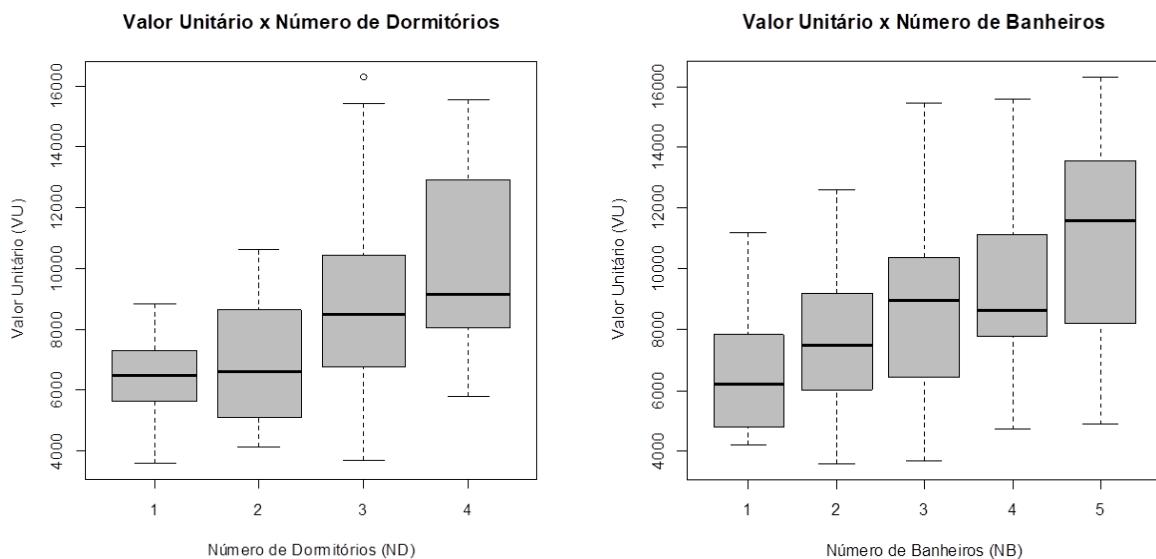


Fonte: o autor

A variável independente número de dormitórios apresentou coeficiente de correlação na ordem de  $r = 0,42$ . Percebe-se que, a medida que aumenta o número de dormitórios de um imóvel, em geral, seu valor unitário tende a aumentar também. Por ser uma variável importante para avaliação de um imóvel e por apresentar crescimento a partir de 2 dormitórios, optou-se por deixá-la na fase de modelagem. O *boxplot* dessa variável pode ser vista na Figura 21 A.

Com relação à variável número de banheiros sociais (NB), percebe-se que, à medida que aumenta o número de banheiros sociais dos imóveis, aumenta também o seu valor unitário. O valor da correlação entre essas duas variáveis é  $r = 0,40$ . Essa correlação pode ser confirmada pelo gráfico *boxplot* da Figura 21 B.

Figura 21 A e B: Gráfico *boxplot* das variáveis dormitórios (ND) e banheiros (NB)



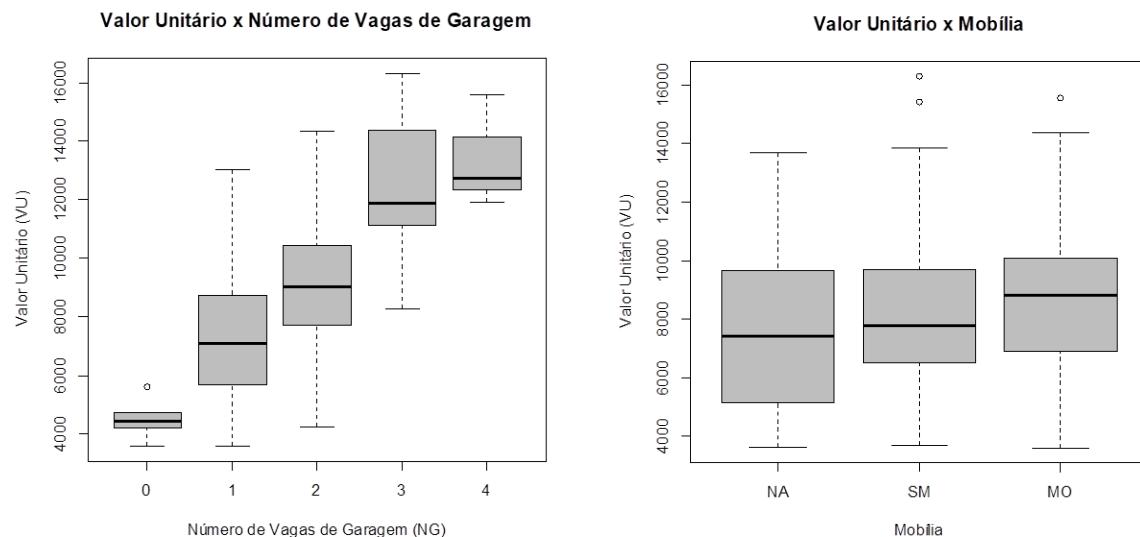
Fonte: o autor

O mesmo ocorre com a variável número de vagas de garagem (NG). Observa-se que os imóveis com mais vagas de garagem, em geral, são mais valorizados que os imóveis com menos vagas de garagens, conforme o gráfico *boxplot* da Figura 22 A, sendo o coeficiente de correlação dessa variável  $r = 0,60$ .

Com relação à variável explicativa mobília no imóvel (MO), decodificada em não mobiliados (NA), semi-mobiliados (SM) e mobiliados (MO), observa-se um aumento do valor unitário dos imóveis, em relação à mediana, à medida que o imóvel vai tendo mais mobília, verificada pelo gráfico *boxplot* mostrado na Figura 22

B. Entretanto, a variável mobília (MO) apresentou coeficiente de correlação  $r = 0,14$ , mostrando pouca influência sobre o valor unitários dos imóveis da amostra. Essa variável será analisada novamente no momento de gerar o modelo de regressão.

Figura 22 A e B: Gráfico *boxplot* das variáveis vagas (NG) e mobília (MO)



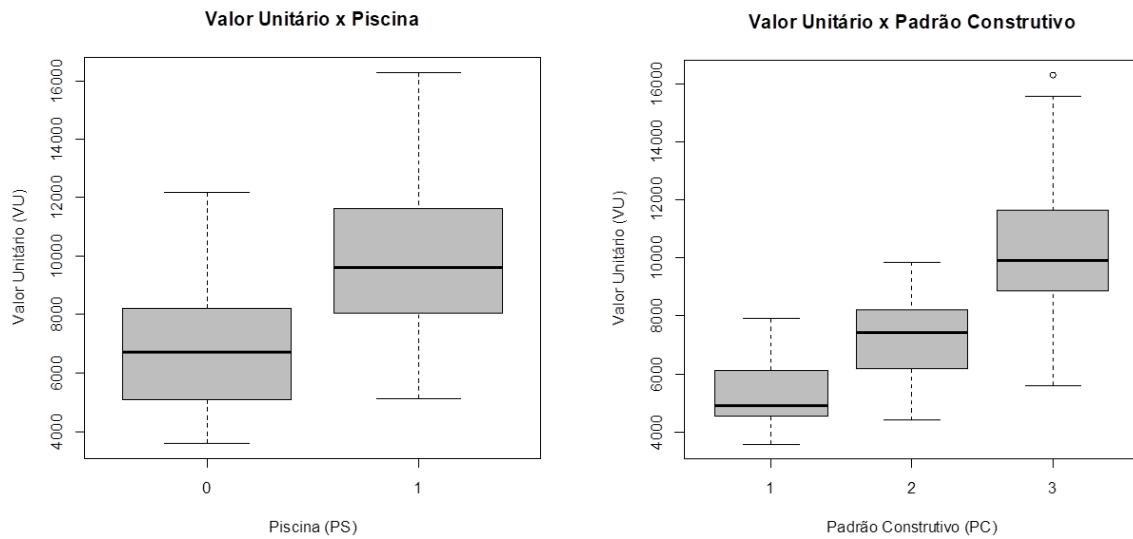
Fonte: o autor

Outra variável analisada é se o condomínio onde se encontra o imóvel coletado tem piscina (PS). Trata-se de uma variável dicotômica em que se codificou 0 quando o condomínio não possuía piscina e 1 quando o condomínio possuía. O coeficiente de correlação entre essa variável e a variável valor unitário é  $r = 0,55$ . Percebe-se, então, que a existência de piscina no condomínio do imóvel faz com que ocorra um aumento do valor unitário dele, conforme visto na Figura 23 A.

Tem-se, também, a variável padrão construtivo (PC), decodificada em baixo, médio e alto, o coeficiente de correlação calculado entre essa variável e a variável valor unitário (VU) foi  $r = 0,75$ . Observa-se que o aumento no padrão de construção torna o valor unitário também maior. Essa relação de dependência pode ser verificada pelo gráfico *boxplot* mostrado na Figura 23 B na sequência.

Por fim, fez-se a análise dos bairros, codificando os dados com 1 se ele se encontre no bairro Agronômica (AG), e 0 caso contrário. Procedeu-se da mesma forma para o bairro Trindade (TR). Nesse caso, se um dado recebe o código 0 para ambas variáveis AG e TR, significa que ela se encontra no bairro Centro.

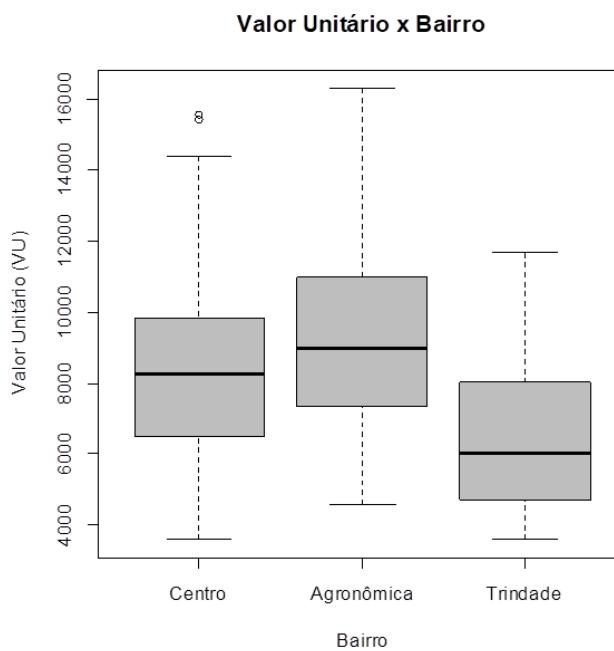
Figura 23 A e B: Gráfico *boxplot* das variáveis piscina (PS) e padrão (PC)



Fonte: o autor

Construiu-se o gráfico *boxplot* do valor unitário dos imóveis em relação a cada um dos bairros em estudo. Observa-se que no Centro e na Agronômica há os imóveis mais valorizados e que o bairro trindade é, dentre estes, o que possui o menor valor unitário de imóveis. O gráfico *boxplot* pode ser visto na Figura 24.

Figura 24: Gráfico *boxplot* representando os bairros em estudo



Fonte: o autor

Desta forma, optou-se por permanecer com as variáveis independentes área privativa (AP), distância à Avenida Beira Mar (DABM), número de dormitórios (ND), número de banheiros (NB), número de vagas de garagem (NG), existência de piscina no condomínio (PS), padrão construtivo (PC) e bairros (AG e TR), utilizando-as na fase de modelagem, verificando sua significância para o modelo e realizando as devidas transformações em suas escalas, caso seja necessário.

## 6.2 MODELO CLÁSSICO DE REGRESSÃO LINEAR MÚLTIPLA

Com o objetivo de encontrar o modelo clássico de regressão que melhor representasse o valor dos imóveis da área em estudo, realizaram-se diversas simulações com transformações nas variáveis independentes, buscando uma modelagem que não violasse os pressupostos básicos do modelo de regressão e que apresentasse significância nos regressores. As variáveis de distância ao polo de valorização foram testadas e desconsideradas aquelas já mencionadas.

O objetivo desse estudo é investigar o uso de modelos de regressão geograficamente ponderada na avaliação em massa de imóveis urbanos, quando comparado com a avaliação de imóveis pela modelagem clássica e espacial.

Nesse sentido, o modelo hedônico geral para esse estudo, considerando que na variável valor unitário será utilizada a transformação logaritmo natural, será:

$$\ln(VU) = f(AP, DPXV, DSBM, DSIG, DCTC, DABM, ND, NB, NS, NG, MO, PS, CH, PC)$$

Verificados os diversos modelos com as transformações necessárias, desconsiderando as variáveis não significativas, eliminando um dado considerado ponto influenciante e um considerado *outlier*, obteve-se a equação de regressão:

$$\ln(VU) = 10,4368 - 0,41225 \times \ln(AP) - 0,10276 \times \ln(DABM) + 0,06686 \times (ND) + 0,04478 \times (NB) + 0,17503 \times (NG) + 0,09764 \times (PS) + 0,20016 \times (PC)$$

Os resultados das estatísticas desse modelo são apresentados na Tabela 10. O modelo foi ajustado com a transformação logaritmo natural nas variáveis valor unitário, área privativa e distância à avenida Beira Mar. Não se considerou, nessa modelagem, a variável indicativa do bairro em que se localizava o imóvel.

Tabela 10: Estatísticas relativas aos parâmetros do modelo OLS<sub>SB</sub>

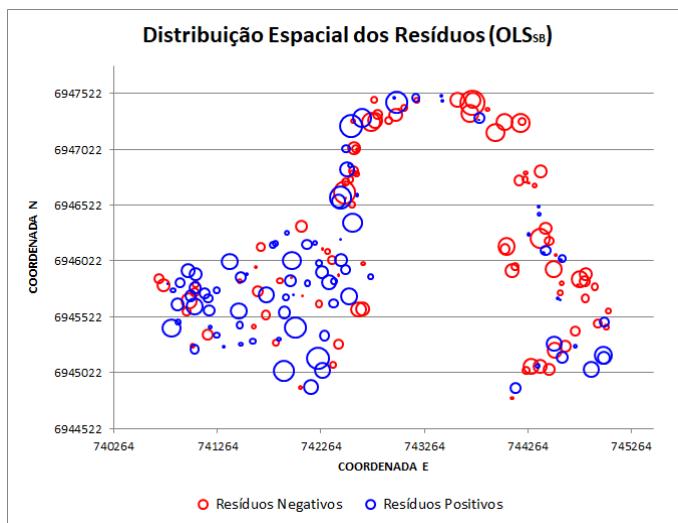
Variável	Coeficientes	Desvio	$t_{\text{calc}}$	Significância
Intercepto	10,4368	0,27067	38,5592	0,00000
In (AP)	-0,41226	0,05842	7,05631	0,00000
In (DABM)	-0,10276	0,01537	6,68776	0,00000
ND	0,06686	0,02285	2,92649	0,00387
NB	0,04478	0,01802	2,48571	0,01384
NG	0,17503	0,02251	7,77449	0,00000
PS	0,09764	0,02788	3,50208	0,00058
PC	0,20016	0,01792	11,1709	0,00000

Fonte: o autor

Os regressores são significativos ao nível  $\alpha = 10,0\%$ , conforme exige a NBR 14.653-2 (2011) para avaliações em que se busca atingir o grau de fundamental III. O modelo clássico apresentado na Tabela 10 apresenta coeficiente de determinação  $R^2$  ajustado de 0,7880, desvio padrão da regressão de 0,1538.

Para se verificar o comportamento espacial dos resíduos desse modelo, construiu-se um diagrama de distribuição espacial, conforme Figura 25.

Figura 25: Distribuição espacial dos resíduos do modelo OLS sem bairro



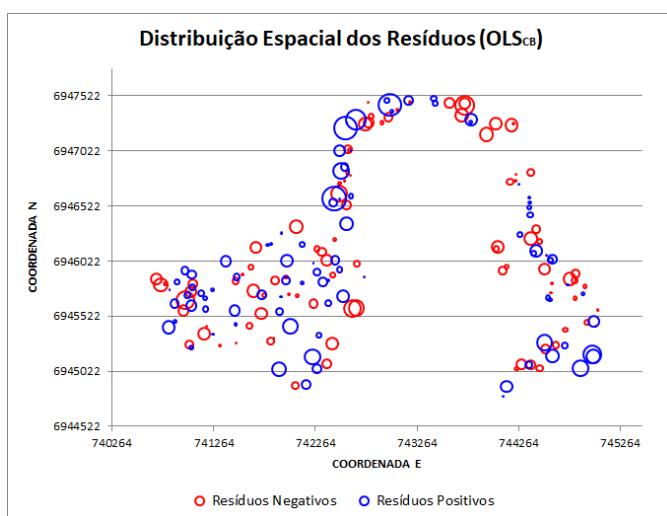
Fonte: o autor

Observa-se, pela Figura 25, que a distribuição espacial dos resíduos sem a presença das variáveis indicativas de bairro apresenta dependência espacial,

evidenciada pela concentração de resíduos positivos no Centro e resíduos negativos na Trindade. No Apêndice 6 é possível consultar as estatísticas da regressão da Tabela 10. O modelo apresenta, ainda, problemas de especificação, multicolinearidade elevada e ausência de variáveis de localização consideradas importantes. Para eliminar esta dependência espacial, incluiu-se no modelo variáveis indicativas dos bairros onde estão localizados os imóveis (vide detalhes no próximo tópico).

Construiu-se, então, o gráfico de distribuição espacial dos resíduos do modelo clássico com a inclusão das variáveis indicativas de bairro ( $OLS_{CB}$ ).

Figura 26: Distribuição espacial dos resíduos do modelo OLS com bairro



Fonte: o autor

Nesse caso, ao se considerar as variáveis indicativas de bairro, percebe-se, pela Figura 26, uma distribuição aleatória dos resíduos positivos e negativos em torno da área em estudo, fazendo com que a dependência espacial que existia anteriormente no modelo fosse fortemente reduzida, resolvendo um possível erro de especificação na forma funcional do modelo clássico de regressão. Testes estatísticos referentes à dependência espacial são apresentados na Tabela 18 (p. 127).

Como esse estudo investiga o uso da regressão GWR para avaliação em massa de imóveis urbanos comparando as estatísticas desta modelagem com as regressões clássica e espacial global, é justo que os modelos obtidos sejam livres de qualquer problema decorrente dos efeitos espaciais. Sendo assim, a modelagem de regressão clássica será realizada considerando as variáveis indicativas de bairro.

### 6.3 MODELO CLÁSSICO COM VARIÁVEIS DE LOCALIZAÇÃO (OLS<sub>CB</sub>)

Para se obter um modelo ideal de comparação, deve-se eliminar todos os efeitos espaciais presentes nos resíduos. O modelo clássico de regressão foi o primeiro modelo de comparação gerado. Os efeitos de dependência espacial foram tratados por meio da inclusão das variáveis de localização indicativas de bairro.

Verificados as diversas transformações possíveis, desconsiderando as variáveis não significativas para o modelo, eliminando dois dados da amostra (um ponto influenciante e um *outlier*), obteve-se a equação de regressão dada por:

$$\ln(VU) = 10,4805 - 0,41074 \times \ln(AP) - 0,10028 \times \ln(DABM) + 0,06671 \times (ND) + 0,03626 \times (NB) + 0,1742 \times (NG) + 0,09969 \times (PS) + 0,2057 \times (PC) - 0,1075 \times (AG) - 0,1076 \times (TR)$$

E, aplicando a definição formal de logaritmo para que se possa isolar a variável explicada valor unitário (VU), obtém-se o modelo de estimação, dado por:

$$VU = e^{10,480 - 0,41074 \times \ln(AP) - 0,10028 \times \ln(DABM) + 0,06671 \times (ND) + 0,03626 \times (NB) + 0,1742 \times (NG) + 0,09969 \times (PS) + 0,2057 \times (PC) - 0,1075 \times (AG) - 0,1076 \times (TR)}$$

Nesse caso, as transformações adotadas para os parâmetros considerados no modelo ajustado foram: logaritmo natural ( $\ln x$ ) para as variáveis valor unitário (VU), área privativa (AP) e distância à avenida Beira Mar (DABM). Permaneceram na escala original ( $x$ ), sem transformações, as variáveis número de dormitórios (ND), número de banheiros (NB), número de vagas de garagem (NG), existência de piscina (PS), padrão construtivo (PC) e os bairros Agronômica e Trindade (AG e TR).

#### 6.3.1 Estatísticas dos Regressores do Modelo

Os valores e as estatísticas referentes à significância dos regressores considerados, ao desvio padrão, à estatística  $t$  e o valor dos coeficientes do modelo estão apresentados na Tabela 11. Foi adotado, conforme sugere a NBR 14.653-2 (2011), como nível de significância máximo para a rejeição da hipótese nula de cada regressor do modelo, por meio do teste bicaudal, o valor  $\alpha = 10,0\%$ , correspondente à estatística  $t_{crit} = 1,6538$ , que é a situação mais desfavorável possível. Observa-se que todos os regressores considerados apresentaram  $t_{calc}$  maior, em módulo, que

$t_{crit}$ , fazendo com que todos sejam considerados significativos para a formação do modelo, confirmado pelas significâncias estarem próximas de zero.

Tabela 11: Estatísticas relativas aos parâmetros do modelo OLS<sub>CB</sub>

Variável	Coeficientes	Desvio	$t_{calc}$	Significância
Intercepto	10,4805	0,25862	40,5254	0,00000
ln (AP)	-0,41074	0,05510	7,45393	0,00000
ln (DABM)	-0,10028	0,01758	5,70322	0,00000
ND	0,06671	0,02152	3,10048	0,00225
NB	0,03626	0,01706	2,12569	0,03491
NG	0,17415	0,02126	8,19078	0,00000
PS	0,09969	0,02690	3,70560	0,00028
PC	0,20571	0,01693	12,1493	0,00000
AG	-0,10746	0,03072	3,49767	0,00059
TR	-0,10758	0,02941	3,65721	0,00034

Fonte: o autor

Os sinais dos regressores confirmam a expectativa do mercado imobiliário local, tornando-os coerentes. É esperado que o aumento de um dormitório, de um banheiro social ou de uma unidade de vaga de garagem provoque um aumento no valor unitário dos imóveis, confirmado pelo sinal positivo. O mesmo ocorre com o aumento do padrão construtivo ou a existência de piscina no condomínio.

Observa-se, ainda, coerência quando se analisa o sinal da variável distância à avenida Beira Mar, considerando que a função é logarítmica. Quanto mais distante o imóvel estiver do polo de valorização, menor o seu valor unitário. O mesmo ocorre com a variável área privativa, onde se espera uma redução do valor do metro quadrado do imóvel quando ocorre um aumento na sua área privativa.

### 6.3.2 Estatísticas do Modelo de Regressão

Na Tabela 12 são apresentadas as estatísticas relacionadas ao modelo de regressão, como o número de variáveis dependentes, de graus de liberdade e de elementos da amostra. É apresentando ainda o valor da estatística  $t$ , do coeficiente de correlação, do coeficiente de determinação e do desvio padrão da regressão.

Tabela 12: Estatísticas relativas ao modelo de regressão OLS<sub>CB</sub>

Estatística	Resultados
Nº de elementos da amostra	188
Nº de variáveis independentes	9
Nº de graus de liberdade	178
Desvio padrão da regressão	0,1448
Coeficiente de correlação	0,9061
Valor <i>t</i> calculado ( <i>t</i> <sub>calc</sub> )	28,580
Valor <i>t</i> tabelado ( <i>t</i> <sub>crit - 1,0%</sub> )	2,6040
Coeficiente de determinação (R <sup>2</sup> )	0,8211
Coeficiente R <sup>2</sup> ajustado (R <sup>2</sup> <sub>ajus</sub> )	0,8121

Fonte: o autor

Pelos resultados da tabela pode-se perceber que o modelo de regressão gerado apresentou coeficiente de correlação na ordem de R = 90,61%, considerando para isso 9 variáveis explicativas e 188 dados amostrais, sendo que dois dados da amostra original foram excluídos (um ponto influenciante e um *outlier*), cuja análise se encontra presente no Tópico 6.3.3 seguinte. Verifica-se, ainda, que o valor do coeficiente de correlação indica que há correlação classificada como fortíssima.

O coeficiente de determinação é uma importante medida descritiva da qualidade do ajuste, entretanto deve-se ter cautela ao utilizá-lo, pois esse coeficiente aumenta à medida que adicionamos novos parâmetros ao modelo de regressão. Com base no coeficiente de determinação ajustado R<sup>2</sup><sub>ajus</sub>, pode-se afirmar que o modelo de regressão obtido é capaz de explicar 81,21% da variação, em relação à média, dos dados de mercado, sendo o restante atribuído a fatores aleatórios e a variáveis não consideradas no modelo. O valor da estatística *t* para o modelo de regressão linear foi calculada para um nível de significância de 1,00%.

Na Tabela 13 é apresentada a análise da variância ou ANOVA, comumente utilizada para descrever um modelo de regressão linear, particionando a variabilidade total dos seus dados em duas componentes: uma devida ao modelo e outra em função do que não é observável, nesse caso relacionada aos resíduos.

A NBR 14.653-2 (2011) sugere que seja realizado o teste *F* de Snedecor para verificar a significância do modelo de regressão ajustado e que, para situação mais desfavorável, o modelo obtido seja significante ao nível de 1,0%.

Tabela 13: Análise da variância do modelo de regressão OLS<sub>CB</sub>

Fonte de Erro	Soma dos Quadrados	Graus de Liberdade	Quadrados Médios	F calculado
Regressão	17,1323	9,0	1,9035	90,780
Residual	3,73250	178	0,0209	
Total	20,8648	187	0,1115	

Fonte: o autor

Verifica-se que, ao nível de significância de 1,0% (NBR 14.653-2, 2011), o valor de  $F_{\text{crit}}$  é igual a 2,508. Sendo assim, como  $F_{\text{calc}}$  é maior que  $F_{\text{crit}}$ , aceita-se a hipótese de existência de regressão com significância próximo de 0,0%.

### 6.3.3 Pressupostos Básicos do Modelo de Regressão

Conforme a NBR 14.653-2, quaisquer que sejam os modelos utilizados para inferir o comportamento do mercado e formação de valores, seus pressupostos devem ser devidamente explicitados e testados, com o objetivo de obter avaliações não tendenciosas, eficientes e consistentes, sendo eles apresentados a seguir:

#### Linearidade

Conforme a NBR 14.653-2 (2011), pode-se examinar a linearidade do modelo por meio da construção de gráficos dos valores observados para a variável dependente versus cada variável independente, com as respectivas transformações. Esses gráficos, que indicam relações lineares, são apresentados no Apêndice 3.

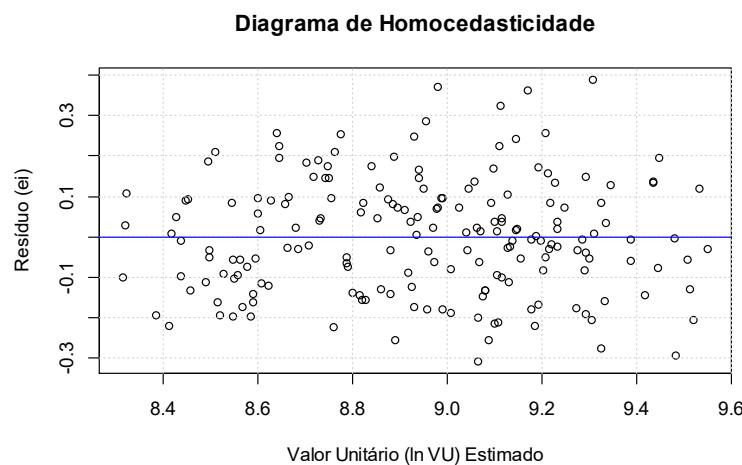
Realizou-se, também, o teste de não-linearidade (quadrados e logaritmos) sob as hipóteses  $H_0$ : a relação é linear e  $H_1$ : a relação não é linear. Obteve-se, para o método dos quadrados,  $p_{\text{calc}} = 0,1480$  e, para o método dos logaritmos,  $p_{\text{calc}} = 0,1067$ . Então, ao nível 5,0%, não se rejeita  $H_0$  e tem-se evidência de linearidade.

#### Homocedasticidade

Um gráfico dos resíduos ( $e_i$ ) versus os valores ajustados pelo modelo de regressão ( $\hat{Y}$ ), apresentando pontos distribuídos aleatoriamente em torno de uma reta horizontal que passa pela origem, sem nenhum padrão definido, é um indicador

favorável à aceitação da hipótese de variância constante para o erro. Na Figura 27 observa-se que o gráfico apresenta distribuição dos resíduos sem tendência definida, de forma totalmente aleatória, indicando possível homocedasticidade.

Figura 27: Gráfico de resíduos x valores estimados pelo modelo



Fonte: o autor

Percebe-se que os resíduos se apresentam aleatoriamente dispostos em torno da reta de referencia  $y = 0$ , que indica se tratar de um modelo homocedástico.

A homocedasticidade se confirma pelos testes de Breusch-Pagan ( $BP = 15,1698$ ) e de Koenker-Basset ( $KB = 17,3558$ ) que apresentou significâncias na ordem de 0,0864 e 0,0513, respectivamente, fazendo aceitar, ao nível de 5,0%, a hipótese nula desejada de que o modelo de regressão em estudo é homocedástico.

## Normalidade

A normalidade dos resíduos é uma suposição essencial para que os resultados do ajuste do modelo de regressão linear sejam confiáveis. Os testes de normalidade são utilizados para verificar se a distribuição de probabilidade associada ao conjunto de dados pode ser aproximada pela distribuição normal.

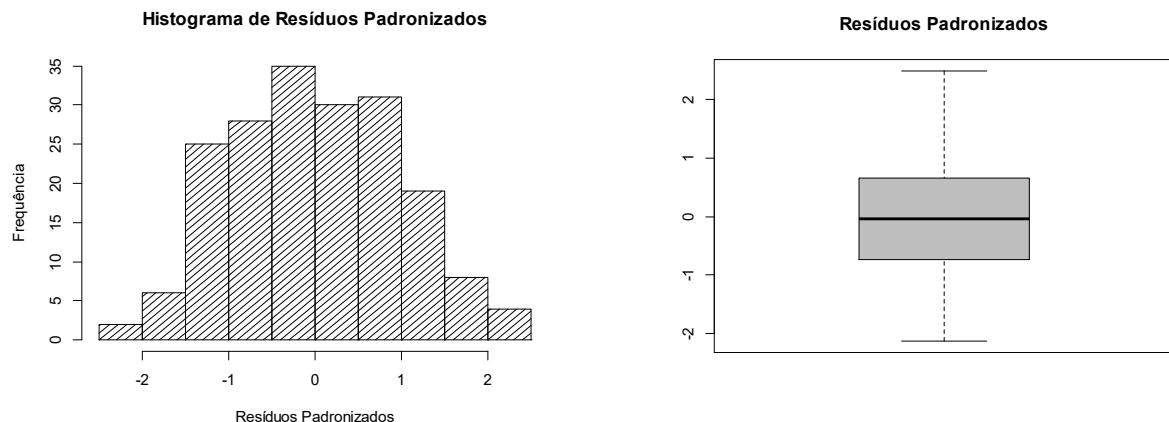
Segundo Dantas (2012), em primeira análise, pode-se fazer a verificação da hipótese de normalidade dos resíduos observando-se o intervalo abrangido pelos resíduos padronizados ( $e_i^*$ ) encontrados dividindo-se cada resíduo ( $e_i$ ) pelo desvio-padrão do modelo, uma vez que, em uma distribuição normal, 68% destes resíduos estão no intervalo [-1,00; +1,00], 90% entre [-1,64; +1,64] e 95% entre [-1,96; +1,96].

Tabela 14: Distribuição dos resíduos normalizados (OLS<sub>CB</sub>)

Intervalo	Distribuição de Gauss	% de Resíduos no Intervalo
[-1,00; +1,00]	68,30%	65,96%
[-1,64; +1,64]	89,90%	92,02%
[-1,96; +1,96]	95,00%	96,28%

Fonte: o autor

Vê-se, pela Tabela 14, que os resultados indicam existência de normalidade nos resíduos, pois os valores obtidos se assemelham aos valores de referência da distribuição normal. Pode-se, também, fazer a análise do histograma e gráfico *boxplot* das Figuras 28 A e B, que apresentam distribuição relativamente simétrica.

Figura 28 A e B: Gráfico *boxplot* e histograma de resíduos padronizados

Fonte: o autor

Verificou-se, também, a normalidade dos resíduos por meio do teste de Kolmogorov-Smirnov cujo valor crítico, para o nível de 5,0%, é  $D_{crit} = 0,1014$ . O cálculo do teste forneceu como maior diferença obtida o valor  $D_{calc} = 0,0367$ . Portanto, segundo o teste de Kolmogorov-Smirnov, a um nível de significância de 5,0%, não se tem evidências para rejeitar a hipótese de normalidade dos resíduos.

Para confirmar a normalidade dos resíduos realizou-se, ainda, o teste de Jarque-Bera, que admite como  $H_0$  que os erros têm distribuição normal e como  $H_1$  que os erros não têm distribuição normal. A estatística do teste forneceu  $JB = 2,7036$  e um  $p$ -valor = 0,2587. Portanto, ao nível de significância de 5,0%, não se pode rejeitar a hipótese nula  $H_0$  e admite-se que há normalidade dos resíduos.

## Multicolinearidade

Conforme Dantas (2012), para verificação desta hipótese deve-se, em primeira mão, analisar a matriz das correlações, que espelha as dependências lineares entre as variáveis explicativas consideradas no modelo de regressão.

Tabela 15: Matriz de correlações das variáveis transformadas

	<b>AP</b>	<b>DABM</b>	<b>ND</b>	<b>NB</b>	<b>NG</b>	<b>PS</b>	<b>PC</b>	<b>AG</b>	<b>TR</b>
<b>AP</b>	1,00								
<b>DABM</b>	0,41	1,00							
<b>ND</b>	0,68	0,30	1,00						
<b>NB</b>	0,69	0,35	0,56	1,00					
<b>NG</b>	0,58	0,19	0,54	0,60	1,00				
<b>OS</b>	0,38	0,34	0,36	0,37	0,46	1,00			
<b>PC</b>	0,27	0,40	0,37	0,34	0,42	0,46	1,00		
<b>AG</b>	0,25	0,55	0,19	0,20	0,13	0,34	0,32	1,00	
<b>TR</b>	0,37	0,58	0,26	0,35	0,20	0,34	0,28	0,36	1,00

Fonte: o autor

Observando-se a matriz de correlações, pode-se concluir que não há indicativos de multicolinearidade forte entre as variáveis explicativas, tendo em vista a não observância de resultados superiores a 0,80, como sugere a NBR 14653-2.

Para confirmar esse resultado, realizou-se o cálculo do fator de inflação da variância (VIF) para todas as variáveis explicativas e verificou-se que o valor máximo ocorreu para variável área privativa com  $VIF_{AP} = 3,17$ , conforme Tabela 16. O cálculo do fator de inflação da variância foi realizado no software R versão 3.5.3.

Tabela 16: Teste do Fator de Inflação da Variância (VIF)

<b>VAR</b>	<b>AP</b>	<b>DABM</b>	<b>ND</b>	<b>NB</b>	<b>NG</b>	<b>PS</b>	<b>PC</b>	<b>AG</b>	<b>TR</b>
VIF	3,17	2,05	1,93	2,75	2,04	1,56	1,61	1,37	1,85

Fonte: o autor

Gujarati *et al.* (2011, p. 348) sugerem que se o fator de inflação da variância (VIF) for superior a 10,0, a variável é tida como altamente colinear. Nesse sentido, todas as variáveis explicativas desse estudo tiveram sua colinearidade aceita.

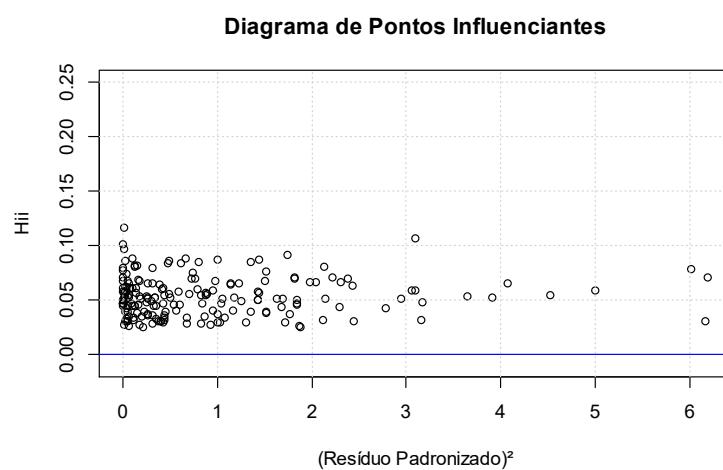
## Verificação de Pontos Influenciantes e *Outliers*

Fez-se, inicialmente, a verificação de pontos influenciantes pelo método da Distância de Cook, que corresponde à variação máxima sofrida pelos coeficientes do modelo quando se retira determinado elemento da amostra. O valor obtido para distância não deve ser maior que o correspondente valor crítico de  $F_{tab}$ , embora seja comumente utilizado 1,0 como distância máxima. Calculou-se a Distância de Cook e obteve-se como maior valor  $d_{calc} = 0,1160$ , que é menor que  $d_{crit} = 1,887$  ( $\alpha = 5,0\%$ ). Conclui-se, então, que não há pontos influenciantes na amostra em estudo.

Para complementar essa análise construiu-se o gráfico de elementos Hii versus o quadrado dos resíduos padronizados. O indicador Hii corresponde aos elementos da diagonal da matriz de previsão e mede a distância da observação para o conjunto das demais observações, sendo que os pontos no canto superior esquerdo do gráfico podem possuir alta influência nos resultados da regressão.

Na amostra de trabalho, o ponto AP\_86 apresentava-se no canto superior esquerdo e tinha impacto sobre os resultados da regressão. Optou-se por retirar esse dado da amostra de trabalho. O gráfico sem esse ponto é visto na Figura 29.

Figura 29: Gráfico de elementos Hii x quadrado dos resíduos padronizados



Fonte: o autor

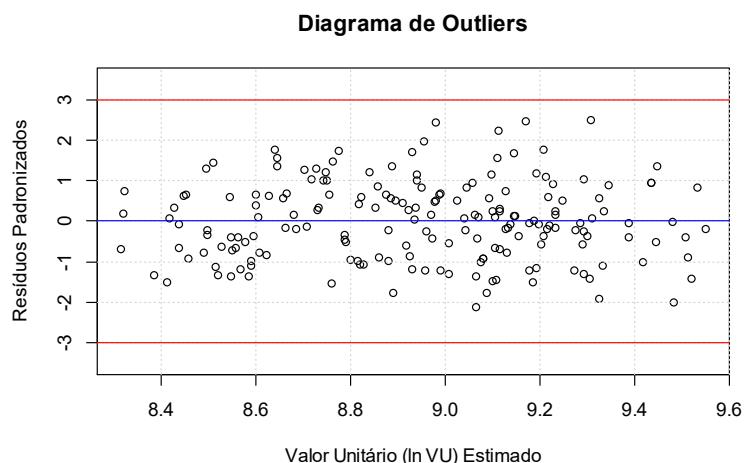
Considera-se *outlier* um dado que contém grande resíduo em relação aos demais que compõem a amostra e podem ser detectados por meio da análise gráfica dos resíduos padronizados ( $e_i$ ) versus os valores ajustados correspondentes ( $\hat{Y}$ ).

Quando em determinada amostra algum elemento apresenta grande dispersão em relação à média, deve-se sanear a amostra, eliminando os dados discrepantes, chamados de *outliers*, sendo comumente utilizado em avaliação de imóveis o intervalo de  $\pm 2,0$  desvios padrões em torno da média, mas esse limite não deve ser rígido, principalmente em amostras com grande quantidade de dados.

Tendo em vista que esse estudo tem relação com a área de engenharia de avaliações, uma amostra contendo 190 dados é vista como grande. Sendo assim, considerou-se *outlier* o ponto que não se encontrava no intervalo de  $\pm 3,0$  desvios padrões. O ponto AP\_115 teve desvio padrão de -3,02 e foi eliminado da amostra.

O gráfico indicativo de *outlier* é visto na Figura 30, em que se observa, após saneamento e exclusão dos dois elementos (AP\_86 e AP\_115), que o mesmo não apresenta nenhum dado além do limite de  $\pm 3,0$  desvios padrões em torno da média.

Figura 30: Gráfico de resíduos padronizados x In (VU) estimado



Fonte: o autor

Desta forma, dois dados foram eliminados da amostra de trabalho, sendo eles: o ponto influenciante AP\_86 (7.556,00) e o *outlier* AP\_115 (6.627,00). O modelo de regressão adotado nesse estudo foi determinado sem esses dois pontos e toda a amostra foi reorganizada para ser utilizada nas etapas seguintes.

Verifica-se, portanto, que o modelo adotado não violou os pressupostos básicos, sendo aprovado em todos os testes realizados, mostrando-se um modelo estatisticamente correto para explicar o mercado imobiliário da área em estudo.

## 6.4 MODELO DE REGRESSÃO ESPACIAL

Verificou-se, no Tópico 6.3 anterior, que a inclusão das variáveis indicativas de bairro (TR e AG) foi capaz de tratar eficazmente a dependência espacial que havia nos resíduos do modelo OLS<sub>SB</sub>. Outra maneira de se realizar esse tratamento é por meio da modelagem por regressão espacial. Nesse tópico será apresentada a modelagem espacial do erro e/ou defasagem (CAR/SAR) como forma de tratar a dependência presente nos resíduos do modelo de regressão linear clássica (OLS<sub>SB</sub>).

Segundo Uberti (2016), para iniciar a análise de regressão espacial deve-se estabelecer a representação matemática que melhor descreve o arranjo espacial dos dados da região, buscando, nesse caso, um modelo que seja capaz de explicar de forma correta a influência que o fator localização pode exercer sobre os dados.

Isso pode ser conseguido testando, inicialmente, os resíduos para autocorrelação espacial por intermédio do valor da estatística I de Moran, usando a matriz de ponderação ou matriz de pesos espaciais. O objetivo é que esta matriz capture a autocorrelação espacial presente nos resíduos do modelo de regressão linear determinado pelo método dos mínimos quadrados ordinários (OLS).

Há diferentes formas de se medir a dependência espacial sendo a estatística I de Moran a mais difundida. Os resultados dessa estatística medem a autocorrelação espacial e podem, também, indicar falta de especificação do modelo.

### 6.4.1 Estimação da Matriz de Pesos Espaciais

Os pesos espaciais  $w_{ij}$  representam o grau de conexão entre as regiões segundo algum critério de proximidade, mostrando a influência da região  $j$  sobre a região  $i$ . Desse modo, a matriz quadrada  $W$  de pesos espaciais é importante por realizar uma ponderação da influência que uma região exerce sobre a outra.

Conforme Almeida (2012), o grau de conexão expresso nas matrizes de pesos espaciais pode ser classificado de acordo com um critério geográfico ou socioeconômico. O critério de conexão geográfica da matriz de pesos espaciais apoia-se na ideia de proximidade que, por sua vez, pode ser definida de acordo com a contiguidade e/ou com a distância geográfica segundo uma determinada métrica.

Segundo Uberti (2016), uma das dificuldades na criação de matrizes de pesos espaciais é a definição de qual a melhor distância ou quantos vizinhos devem

ser utilizados, para capturar a maior parte da autocorrelação espacial. Nesse sentido, a recomendação é testar várias matrizes de pesos para ter certeza de qual delas é capaz de capturar a maior parte da autocorrelação espacial.

Sendo assim, foram testadas diversas matrizes de ponderação espacial usando como critérios geográficos a distância (m) e a vizinhança (k) com os respectivos testes globais de Moran. Em ambos os critérios a matriz de pesos foi construída com base no inverso da distância. O valor dos testes globais de Moran foram aplicados sobre os resíduos do valor unitário dos imóveis em estudo.

Tabela 17: Teste global de Moran para matrizes de ponderação espacial

Vizinhos (k)	Moran's I	Vizinhos (k)	Moran's I	Distância (m)	Moran's I	Distância (m)	Moran's I
4	0,1310	10	0,1020	300	0,1016	550	0,1309
5	0,1054	10	0,1058	350	0,1144	575	0,1302
6	0,0990	12	0,1123	400	0,1157	600	0,1234
7	0,0926	13	0,1137	450	0,1234	650	0,1248
8	0,1016	14	0,1043	500	0,1257	700	0,1210
9	0,1017	15	0,1028	<b>525</b>	<b>0,1313</b>	750	0,1090

Fonte: o autor

O valor I de Moran tem um valor esperado de  $E(I) = -1/(n - 1)$ , isto é, o valor que seria obtido se não houvesse padrão espacial dos dados. O valor calculado de I deve ser igual ao seu valor esperado, dentro dos limites de significância estatística, se  $y_i$  for independente dos valores das regiões vizinhas (ALMEIDA, 2012, p. 106).

Tem-se, nesse estudo,  $n = 188$  regiões, o que forneceria  $E(I) = -0,005$ . Pode-se observar, pela Tabela 17, que todas as distâncias e número de vizinhos testados apresentaram valores para a estatística I de Moran diferentes e maiores que o esperado, o que indica a presença de autocorrelação positiva nos dados.

O teste global I de Moran presta-se a uma estatística cuja hipótese nula é a de independência espacial e nesse caso seu valor seria zero. Valores positivos e portanto entre  $0 < I \leq 1$  indicam correlações diretas e valores negativos e portanto entre  $-1 \leq I \leq 0$  indicam correlações inversas. É importante que, após calculada, verifique-se a significância da estatística por meio do teste de pseudossignificância em que são geradas diferentes permutações de valores de atributos dos imóveis.

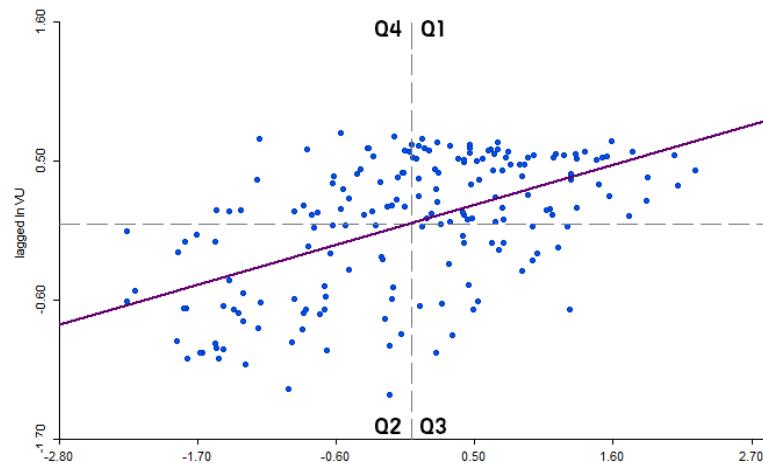
Verifica-se que a largura de banda  $d = 525$  m é a que apresenta o maior valor para o Índice de Moran ( $I = 0,1313$ ) e será essa a distância adotada nesse estudo com o objetivo de tentar capturar a maior parte da autocorrelação espacial.

Realizou-se o teste de pseudosignificância com 999 permutações, por meio do software GeoDa e obteve-se um nível de significância de 1,0%, o que indica que há fortes indícios para se rejeitar a hipótese nula de independência espacial.

Como forma adicional de verificar a dependência espacial, há o diagrama de espalhamento de Moran, instrumento construído tendo como base os valores normalizados, ou seja, no contexto desse estudo teríamos a média subtraída dos valores unitários dos imóveis e dividida pelo desvio padrão da amostra.

Esse diagrama permite avaliar como se dá a variabilidade espacial comparando os valores unitários normalizados, em determinada área, com a média dos seus vizinhos, construindo um gráfico de  $\ln(VU)$  (valores normalizados) por  $w(\ln VU)$  (média dos vizinhos). O diagrama de Moran é mostrado na Figura 31.

Figura 31: Diagrama de espalhamento de Moran para a variável  $\ln(VU)$



Fonte: o autor

O diagrama de espalhamento de Moran anterior foi produzido no software GeoDa e corrobora com as conclusões anteriores. Os valores dos quadrantes Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas) indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores unitários de imóveis considerados semelhantes.

Como se pode ver, a maioria dos imóveis da amostra estão concentrados nos quadrantes Q1 e Q2 e apresentam associação espacial positiva, evidenciando

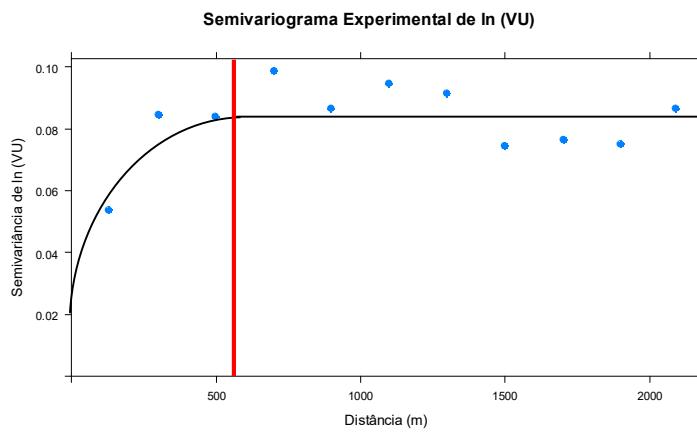
regiões cujos valores unitários estão acima da média, mas rodeada por vizinhos de valores unitários altos (Q1), bem como regiões cujos valores unitários estão abaixo da média, porém rodeado de vizinhos com valores unitários baixos (Q2).

Segundo Câmara *et al.* (2002), a hipótese implícita no cálculo do índice de Moran é a estacionariedade de primeira e de segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. A estacionariedade existe quando a autocorrelação espacial apresenta comportamento homogêneo na região de estudos. Quando existir não-estacionariedade de primeira ordem, o que chamamos de tendência, os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice.

Como poderá ser visto, os dados desse estudo não se mostraram estacionários espacialmente, apresentando padrões distintos ao longo da área em estudo. Esse fenômeno será tratado por meio de regressões geograficamente ponderadas.

Testou-se também o uso de semivariograma experimental para determinação da matriz de pesos espaciais. O semivariograma foi construído sobre o logaritmo dos valores unitários e ajustado no software R na sua versão 3.5.3. Utilizou-se na construção do semivariograma o pacote *gstat* e suas respectivas bibliotecas geoestatísticas. O semivariograma construído pode ser observado na Figura 32.

Figura 32: Semivariograma experimental de ln do valor unitário observado



Fonte: o autor

Os resultados acima concordam com a hipótese de que a distância  $d = 525,00$  é a que captura melhor a autocorrelação espacial dos dados dos imóveis utilizados nesse estudo. Desta forma, pode-se concluir que as observações distantes em até 525 metros exercem maior influência entre si e esta influência diminui a

medida que a distância aumenta, de modo que aquelas que se encontram separadas por distâncias maiores quase não apresentam influência entre si.

#### 6.4.2 Indicador de Autocorrelação Espacial: Multiplicador de Lagrange

Definida a matriz de pesos espaciais deve-se seguir para definição de qual o melhor modelo a se ajustar. O resultado obtido no teste I de Moran indicou que há autocorrelação espacial nos resíduos do modelo de regressão clássico ajustado (sem os bairros), entretanto não apontou qual o tipo de autocorrelação espacial é predominante: se no termo da variável dependente ou se no termo do erro.

Anselin (2005) assinala que para se resolver o problema de identificação dos modelos de regressão espacial é recomendado que se utilize testes focados, pois estes são capazes de identificar a autocorrelação espacial e determinar se ela ocorre predominantemente na variável dependente ou nos termos do erro.

Nesta etapa do estudo, fez-se o cálculo do índice Moran's I para modelagem espacial considerando a inclusão e a não inclusão das variáveis dicotômicas bairros (AG e TR). Realizou-se, também, os testes dos multiplicadores de Lagrange (*LM*) clássico e robusto sobre a matriz de ponderação espacial  $d = 525$  m, calculados por meio do software GeoDa. Os valores obtidos são mostrados na Tabela 18.

Tabela 18: Teste dos Multiplicadores de Lagrange (*LM*) com e sem bairro

<b>Estatística</b>	<b>SEM Variável Bairro</b>		<b>COM Variável Bairro</b>	
	Valor	p-Valor	Valor	p-Valor
Moran's I (erro)	6,9834 (0,1313)	0,000	1,6743 (0,0101)	0,094
<i>LM</i> (lag)	7,3667	0,006	0,1123	0,738
<i>LM</i> ROBUSTO (lag)	0,1318	0,716	0,0307	0,860
<i>LM</i> (erro)	16,348	0,002	0,0969	0,756
<i>LM</i> ROBUSTO (erro)	9,1136	0,000	0,0153	0,902

Fonte: o autor

Percebe-se que, ao não se considerar as variáveis indicativas de bairro, o teste de Lagrange detecta a presença de autocorrelação espacial na variável

dependente ( $LM_{LAG} = 0,60\%$ ) e nos termos do erro ( $LM_{ERR} = 0,20\%$ ). Já quando se considera as variáveis indicativas de bairro, o teste de Lagrange indica que, ao nível de significância de 5,0%, toda a autocorrelação espacial dos dados foi resolvida.

Os resultados da Tabela 18 indicam, portanto, que a inclusão dessas variáveis na modelagem foi capaz de reduzir totalmente a dependência espacial que existia tanto na variável dependente quanto nos resíduos da regressão clássica.

Os testes de especificação dos Multiplicadores de Lagrange mostraram que, ao nível de significância de 5,0% e sem considerar as variáveis indicativas de bairro, tanto o  $LM$  (lag) quanto o  $LM$  (erro) se tornaram significativos, evidenciando dependência espacial na variável dependente e nos termos de erro do modelo de regressão clássico. Nesse caso, o critério robusto definirá a modelagem que melhor se ajusta aos dados, sendo, assim, apenas o  $LM_{ROBUSTO}$  (erro) significativo.

Como forma de comparação entre o modelo do erro espacial ( $CAR$ ) e do modelo da defasagem espacial ( $SAR$ ) é apresentado na Tabela 19 o valor obtido para a estatística do critério de informação de Akaike ( $AIC$ ) e do critério de Schwarz ( $BIC$ ) para a modelagem sem a presença da variável dicotômica indicativa de bairro.

Tabela 19: Estatísticas dos modelos do erro e da defasagem espacial

Estatística	$SAR_{SB}$	$CAR_{SB}$
Critério de Akaike ( $AIC$ )	– 167,679	– 178,696
Critério Schwarz ( $BIC$ )	– 138,552	– 152,805

Fonte: o autor

Comprova-se, assim, que entre as duas modelagens, a regressão do erro espacial ( $CAR$ ) apresentou um valor menor no critério de informação de Akaike e no critério de Schwarz, indicando maior aderência. Os resultados da tabela sugerem que o modelo do erro espacial (*Spatial Error*) foi capaz de capturar de forma mais eficiente a relação espacial que existe entre as variáveis em estudo.

Sendo assim, para esse estudo, será adotada a modelagem do erro espacial (*Spatial Error Model – CAR*) sem considerar as variáveis indicativas de bairro. Busca-se, assim, mostrar que a modelagem espacial é uma maneira alternativa para se tratar a dependência espacial que existe nos resíduos da regressão clássica.

#### **6.4.3 Estimação do Modelo de Regressão Espacial do Erro (CAR)**

Definido o modelo do erro espacial (CAR<sub>SB</sub>) como o modelo de regressão espacial que melhor se ajustaria aos dados dos imóveis da área em estudo e estabelecida a matriz de pesos espaciais ( $d = 525$  m), é o momento de calcular os coeficientes do modelo e as respectivas significâncias por meio da estatística  $z$ .

Segundo Trivelloni (2005), a estatística  $z$  corresponde ao equivalente para a regressão por máxima verossimilhança ao valor  $t$  de Student para o método de mínimos quadrados. As probabilidades indicam o grau de significância de cada variável de forma análoga a da regressão por mínimos quadrados.

No modelo de regressão espacial do erro a variável lambda ( $\lambda$ ) representa o termo de autocorrelação espacial do erro e se mostra significativa ao nível de 5,0%.

Observa-se, ainda, que todas as variáveis do modelo de regressão espacial do erro se mostraram significativas ao nível de 5,0%. Obteve-se a equação:

$$\ln(VU) = 10,6470 - 0,43255 \times \ln(AP) - 0,11770 \times \ln(DABM) + 0,05794 \times (ND) + 0,03583 \times (NB) + 0,1852 \times (NG) + 0,09129 \times (PS) + 0,2017 \times (PC) + 0,5691 \times (Wu)$$

E, aplicando a definição formal de logaritmo para que se possa isolar a variável explicada VU, obtendo o modelo para a variável dependente, tem-se:

$$VU = e^{10,6470 - 0,43255 \times \ln(AP) - 0,11770 \times \ln(DABM) + 0,05794 \times (ND) + 0,03583 \times (NB) + 0,1852 \times (NG) + 0,09129 \times (PS) + 0,2017 \times (PC) + 0,5691 \times (Wu)}$$

Nesse caso, como se pode observar pelo modelo ajustado, as transformações adotadas são iguais àquelas adotadas no modelo clássico de regressão.

#### **6.4.4 Estatísticas dos Regressores do Modelo**

Os valores e estatísticas referentes à significância dos regressores considerados, ao desvio padrão, à estatística  $z$  e o valor dos coeficientes do modelo estão apresentados na Tabela 20. Foi adotado como nível de significância máximo para a rejeição da hipótese nula de cada regressor o valor  $\alpha = 10,0\%$ , conforme exige a NBR 14.653-2 (2011) para que se atinja o Grau III de fundamentação.

Tabela 20: Estatísticas relativas aos parâmetros do modelo CAR<sub>SB</sub>

Variável	Coeficientes	Desvio	$z_{\text{calc}}$	Significância
Intercepto	10,6470	0,27877	38,1924	0,00000
In (AP)	-0,43255	0,05537	7,81141	0,00000
In (DABM)	-0,11770	0,02200	5,35113	0,00000
ND	0,05794	0,02097	2,76309	0,00573
NB	0,03583	0,01688	2,12263	0,03379
NG	0,18520	0,02075	8,92755	0,00000
PS	0,09129	0,02631	3,47056	0,00052
PC	0,20171	0,01636	12,3279	0,00000
$\lambda$ (Lâmbda)	0,56916	0,12452	4,57072	0,00000

Fonte: o autor

Observa-se que todos os regressores são significativos e importantes para a formação do modelo, confirmado pelas significâncias serem baixas e estarem próximas de zero. Os sinais dos regressores do modelo de regressão espacial do erro se apresentaram iguais aos sinais do modelo clássico de regressão linear e confirmam a expectativa do mercado imobiliário local, sendo, portanto, coerentes.

#### 6.4.5 Estatísticas do Modelo de Regressão

Na Tabela 21 é apresentado as estatísticas relacionadas ao modelo de regressão, como o total de variáveis, o coeficiente de determinação e desvio padrão.

Tabela 21: Estatísticas relativas ao modelo de regressão CAR<sub>SB</sub>

Estatística	Resultados
Nº de elementos da amostra	188
Nº de variáveis independentes	8
Nº de graus de liberdade	180
Desvio padrão da regressão	0,1425
Coeficiente de correlação	0,9038
Coeficiente de determinação ( $R^2$ )	0,8170

Fonte: o autor

Destaca-se que, com base no coeficiente de determinação  $R^2$ , pode-se afirmar que o modelo de regressão obtido é capaz de explicar 81,70% da variação, em relação à média, dos dados de mercado, sendo o restante atribuído a variáveis não consideradas no modelo. Verifica-se, ainda, que o valor do coeficiente de correlação do modelo indica que há correlação classificada como fortíssima.

Por fim, o teste de Breusch-Pagan para os resíduos do modelo de regressão espacial do erro indicaram que não há problemas de heterocedasticidade. Com um valor de  $BP = 6,57$  e um  $p$ -valor = 0,4740 pode-se afirmar, ao nível de significância de 5,0%, que o modelo do erro espacial autorregressivo é homocedástico.

## 6.5 MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA

Conforme Uberti (2016), a regressão geograficamente ponderada (GWR) é uma técnica que estende o método tradicional da regressão, permitindo que os parâmetros sejam estimados localmente, para que os coeficientes do modelo, em vez de serem estimados globais, sejam específicos para um ponto de localização  $i$ .

Como visto, para se realizar a regressão geograficamente ponderada, é necessário que se defina a função de ponderação espacial (*kernel*) a ser aplicada, conforme Tabela 2 (Página 68), podendo ser do tipo gaussiana ou biquadrática, e a largura de banda (*bandwidth*) considerada, podendo ser do tipo fixa ou variável.

Para escolha da melhor função de ponderação, se fixa ou variável, e melhor largura de banda, foi utilizado, conforme descrito na metodologia, o *software* GWR4 de autoria de Fotheringham *et al.* (2002) e realizadas diversas simulações, separadas por critérios de escolha definido pelo usuário na interface do *software* GWR4.

O *software* permite que se escolha o tipo de função de ponderação (*kernel*) entre os quatro citados anteriormente, a largura de banda e o critério de parada, estando disponível como critério de parada aquela iteração que fornecer o menor valor AIC, ou menor valor de AICc, ou menor valor de CV (*cross validation*).

Interessante citar que o *software* permite marcar a caixa de seleção da melhor largura de banda (*bandwidth*). Ele então gera diversos modelos para diversas larguras de banda a partir da função de ponderação (*kernel*) selecionada e apresenta, por fim, aquele que possui o menor valor para o critério selecionado.

Para especificação da *bandwidth*, o valor deve ser uma métrica de distância no caso de *kernel* fixo, enquanto o valor deve ser o número de vizinhos mais próximos a serem usados para o processo de ajuste em o caso do *kernel* variável.

Realizou-se, então, uma simulação para cada uma das quatro funções de ponderação espacial, com o critério de seleção automática da melhor largura de banda, para cada um dos três critérios de escolha do melhor modelo. Os resultados podem ser vistos nas Tabelas 22, 23 e 24 mostradas na sequência.

Foram gerados 12 modelos sendo que o escolhido seria aquele que fornecesse o menor valor no critério de informação de Akaike (AIC), parâmetro que será utilizado para se fazer a comparação entre os modelos gerados nesse estudo. Observa-se que a regressão GWR, por ser uma modelagem local, já incorpora em seus cálculos a posição geográfica dos dados, sem necessidade de indicar o bairro.

Tabela 22: Escolha pelo Critério de Informação de Akaike (*AIC*)

<b>Função Ponderação</b>	<b><i>AIC</i></b>	<b>Bandwidth</b>
<b>Gaussiana Fixa</b>	<b>– 207,57</b>	<b>278,42</b>
Gaussiana Variável	– 183,70	56
Biquadrática Fixa	– 194,42	1.084,56
Biquadrática Variável	– 190,63	131

Fonte: o autor

Tabela 23: Escolha pelo Critério de Akaike Corrigido (*AICc*)

<b>Função Ponderação</b>	<b><i>AIC</i></b>	<b>Bandwidth</b>
Gaussiana Fixa	– 192,33	987,76
Gaussiana Variável	– 183,04	63
Biquadrática Fixa	– 191,22	2.812,44
Biquadrática Variável	– 190,03	143

Fonte: o autor

Tabela 24: Escolha pelo Critério de Validação Cruzada (*CV*)

<b>Função Ponderação</b>	<b><i>AIC</i></b>	<b>Bandwidth</b>
Gaussiana Fixa	– 193,44	886,02
Gaussiana Variável	– 183,70	56
Biquadrática Fixa	– 192,04	2.586,96
Biquadrática Variável	– 190,52	133

Fonte: o autor

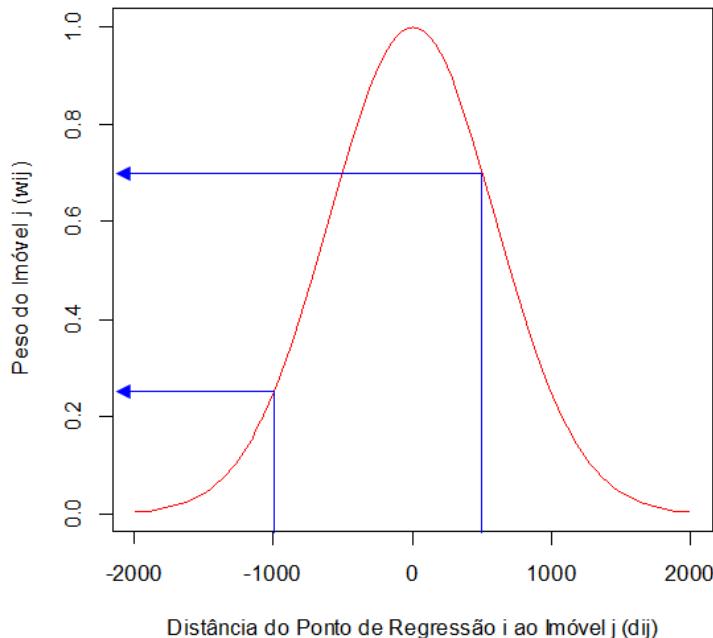
Destaca-se que, nas simulações, geraram-se modelos considerando todas as variáveis como sendo locais. O processo de análise de quais variáveis locais poderiam ser tratadas como globais será realizado na etapa seguinte.

Nesse caso, o modelo de trabalho selecionado para esse estudo foi o que apresentou  $AIC = -207,57$  da Tabela 22. Esse modelo não considera as variáveis indicativas de bairro, possuindo, portanto, sete variáveis explicativas, utiliza a função de ponderação espacial Gaussiana Fixa e a largura de banda que forneceu o menor AIC foi 278,42 m. Esse parâmetro definirá a suavização da curva gaussiana e determinará a influência que um imóvel  $j$  terá no ajuste de uma observação  $i$ .

Para ilustrar essa situação, construiu-se a curva gaussiana considerando a Equação (1) da Tabela 2 e uma largura de banda  $\beta = 800,00$ , mostrada a seguir.

Figura 33: Curva gaussiana fixa para a *bandwidth* hipotética

**Função Gaussiana Fixa**



Fonte: o autor

Observa-se, por exemplo, que se um imóvel  $j$  está a 1000 metros de distância de um ponto de regressão  $i$ , ele terá um peso aproximado de  $w_{ij} = 0,25$  no ajuste de  $i$ . Já um imóvel  $j$  terá um peso  $w_{ij} = 0,70$  quando estiver a 500 metros de distância de um ponto de regressão  $i$ . Por fim, um imóvel  $j$ , que estiver, por exemplo, a 2000 metros de distância de  $i$ , contribuirá com aproximadamente  $w_{ij} = 0,03$ .

#### 6.5.1 Teste DIFF de Estacionariedade dos Parâmetros do Modelo

Como visto, há ainda a possibilidade de, entre as sete variáveis explicativas consideradas nesse estudo, ter alguma que apresente estacionariedade espacial, ou seja, não tenha uma variabilidade espacial relevante ao longo do plano geográfico.

Segundo Leung *et al.* (2000), é importante saber se os parâmetros em um modelo de regressão geograficamente ponderada variam ou não significativamente. Para o autor, deve-se fazer a pergunta: o conjunto de estimativas dos parâmetros locais  $\beta_{ij}$  ( $i = 1, 2, 3, \dots, n$ ) exibe variação espacial significativa na região de estudo?

Quando isso não ocorre, é aconselhável que essa variável seja tratada como global e se utilize um modelo misto, conforme visto na Equação 4.21 do texto.

Em geral, a variabilidade geográfica para cada coeficiente pode ser testada por comparação de modelos. Para testar a variabilidade geográfica do  $k$ -ésimo coeficiente variável, uma comparação de modelo é realizada entre o GWR ajustado e um modelo de teste no qual apenas o  $k$ -ésimo coeficiente é fixo, enquanto outros coeficientes são mantidos como no modelo geograficamente ponderado ajustado.

Nesse caso, se o GWR ajustado for melhor que o modelo GWR de teste, utilizando como critério de comparação o AICc, pode-se concluir que  $k$ -ésimo coeficiente certamente varia no espaço. Essa rotina de teste repete essa comparação entre modelos para cada coeficiente geograficamente variável.

Na Tabela 25 pode ser visto o teste que mostra nas linhas os supostos termos locais e na última coluna o Critério DIFF, que mostra a diferença no indicador de comparação entre o modelo GWR ajustado e o modelo GWR de teste.

Tabela 25: Teste de estacionariedade dos parâmetros do modelo GWR<sub>SB</sub>

Variável	F	DOF	Teste F	Critério DIFF
Intercepto	660,72	6,122	103,87	– 681,00
AP	180,52	7,942	103,87	– 490,75
DABM	1045,4	3,873	103,87	– 685,68
ND	1,61	8,060	103,87	– 6,0723
NB	1,80	8,495	103,87	– 8,8288
NG	1,81	7,977	103,87	– 8,5496
PS	0,99	8,023	103,87	2,2167
PC	2,16	8,246	103,87	– 13,272

Fonte: o autor

Para esse teste, conforme descrito por Leung *et al.* (2000), um Critério DIFF positivo, especialmente maior ou igual a dois, sugere que o termo local é melhor para ser assumido como global. No teste, a maioria das variáveis analisadas apresentaram valores no Critério DIFF menores que dois e negativas, fazendo com que essas variáveis sejam assumidas como locais no modelo de regressão.

Pode-se ver, entretanto, que a variável existência de piscina (PS) apresenta Critério DIFF positivo, o que sugere que essa variável pode ser tratada como global,

caso seja interessante. Simulações indicaram que ao se considerar a variável existência de piscina no condomínio (PS) como global, o critério de informação AIC passava de – 207,57 para – 209,78, resultando em um incremento de 2,21 unidades.

É importante ressaltar que não seria errado utiliza-la como local, entretanto ocorre uma melhora no ajuste do modelo quando ela é tratada como global.

### **6.5.2 Identificação de Coeficientes Constantes no Modelo**

Durante a modelagem por regressão geograficamente ponderada, a variável dicotômica existência de piscina (PS) frequentemente apresentava valor do critério AIC menor quando era tratada como variável global (estacionária).

O teste Critério DIFF, mostrando anteriormente, já apontava essa situação. Realizou-se, então, o test  $t$  para essa variável. Na Tabela 26 é apresentado a variável cujo valor calculado para a estatística  $t$  foi superior ao valor do  $t$  crítico, se mostrando significativa. Portanto, ao nível de significância de 5,0%, admite-se que a variável piscina (PS) apresenta variação em função da posição geográfica.

Tabela 26: Estatística relativa à variável PS do modelo GWR<sub>SB</sub>

Variável	Coeficientes	Desvio	$t_{calc}$	Significância
Piscina (PS)	0,09732	0,02892	3,36478	0,00000

Fonte: o autor

Verifica-se, portanto, que o teste  $t$  acima mostrou que a variável PS poderia ser tratada como local, pois se aceitou a hipótese alternativa. Entretanto, apesar de significativa, as simulações continuavam a melhorar o ajuste nos momentos em que essa variável era excluída da modelagem local, e passava-se a trata-la como global, confirmando o que se observou ao realizar o teste do Critério DIFF, fazendo, portanto, com que ela seja considerada estacionária ao longo desse estudo.

### **6.5.3 Modelo Misto de Regressão Geograficamente Ponderada**

Como visto, uma das sete variáveis consideradas apresentou estacionariedade espacial e foi tratada como global, enquanto as demais foram tratadas como

locais. Nesse sentido, o modelo matemático apresentado a seguir, é composto por seis variáveis explicativas locais (AP, DABM, ND, NB, NG, PC) onde se utilizou o coeficiente médio dado pela Tabela 27, e uma variável explicativa global (PS).

$$\ln(VU) = 10,7317 - 0,51700 \times \ln(AP) - 0,08873 \times \ln(DABM) + 0,06497 \times (ND) + 0,06173 \times (NB) + 0,1907 \times (NG) + 0,09732 \times (PS) + 0,2073 \times (PC)$$

Importante mencionar, novamente, que se tem, aqui, coeficientes médios para as variáveis locais e coeficientes fixos para a variável global (PS). Dessa forma, isolando a variável explicada VU, obtém-se o modelo de estimação dado por:

$$VU = e^{10,7317 - 0,51700 \times \ln(AP) - 0,08873 \times \ln(DABM) + 0,06497 \times (ND) + 0,06173 \times (NB) + 0,1907 \times (NG) + 0,09732 \times (PS) + 0,2073 \times (PC)}$$

No modelo de regressão geograficamente ponderada (GWR) anterior, como se pode observar, permaneceram as transformações adotadas na modelagem de regressão clássica e de regressão do erro espacial demonstrados anteriormente.

#### 6.5.4 Estatísticas dos Regressores do Modelo

Como na regressão GWR é gerado um modelo para cada observação, as estatísticas relativas à estimativa dos parâmetros, ao teste  $t$ , ao desvio padrão e ao coeficiente de determinação são fornecidas para cada um dos regressores de cada um dos 188 modelos, gerando um relatório extenso. No Apêndice 04 desse estudo consta as estatísticas mencionadas para 50 dados de mercado da amostra.

Na Tabela 27 são listados os coeficientes médio e mediano para as variáveis explicativas do modelo GWR, classificadas em variável local ou global.

Tabela 27: Estatísticas relativas aos parâmetros do modelo GWR<sub>SB</sub>

Variável	Coeficiente Médio	Coeficiente Mediano	Desvio	Tipo
Intercepto	10,73166	10,73542	1,14425	Local
ln (AP)	-0,51700	-0,50740	0,20169	Local
ln (DABM)	-0,08873	-0,10434	0,13173	Local
ND	0,06497	0,06628	0,05816	Local

NB	0,06173	0,06011	0,06660	Local
NG	0,19071	0,19394	0,06174	Local
PS	0,09732	0,09732	0,02890	Global
PC	0,20735	0,22325	0,05966	Local

Fonte: o autor

Os sinais dos regressores do modelo de regressão geograficamente ponderada se apresentaram iguais aos sinais do modelo clássico de regressão linear e confirmam a expectativa do mercado imobiliário local, sendo, portanto, coerentes.

#### 6.5.5 Estatísticas do Modelo de Regressão

Na Tabela 28 é apresentado as estatísticas relacionadas ao modelo de regressão, como o coeficiente de validação, de determinação e seu desvio padrão.

Tabela 28: Estatísticas relativas ao modelo de regressão GWR<sub>SB</sub>

Estatística	Resultados
Desvio padrão da regressão	0,1303
Coef. de Validação Cruzada ( $CV$ )	0,0252
Largura de Banda Ótima	278,42
Coeficiente de determinação ( $R^2$ )	0,9239
Coeficiente $R^2$ ajustado ( $R^2_{ajust}$ )	0,8460

Fonte: o autor

Destaca-se que, com base no coeficiente de determinação  $R^2$  ajustado, pode-se afirmar que o modelo de regressão foi capaz de explicar 84,60% da variação, em relação à média, dos dados de mercado, sendo o restante atribuído a variáveis não consideradas no modelo. A regressão GWR apresentou, ainda, um desvio padrão de 0,1303, o mais baixo dentre as três modelagens realizadas.

Por fim, o modelo GWR gerado considerou como largura de banda ótima o valor  $\beta = 278,42$  m e como função de ponderação ideal (*kernel*) a Gaussiana Fixa.

## 6.6 PERFORMANCE DOS MODELOS DE REGRESSÃO

No tópico seguinte será apresentada uma comparação sobre a qualidade dos modelos ajustados, tendo como base o valor do critério de informação de Akaike (*AIC*) e o critério log likelihood (*LIK*), bem como o desempenho das regressões usando uma amostra de validação para quantificar os indicadores COD e PRD.

### 6.6.1 Qualidade do Ajuste dos Modelos

É comum o uso do critério de informação de Akaike (*AIC*), que incorpora em seus cálculos o valor da função de máxima verossimilhança, para a seleção de modelos, sendo que o melhor modelo é aquele que apresenta o menor valor de *AIC*.

Utilizou-se também o critério de informação Bayesiano (*BIC*) para avaliar os modelos obtidos. Esse critério de informação baseia-se, em parte, sobre a função de probabilidade e está estreitamente relacionado com o critério de Akaike (*AIC*).

Nesse estudo o valor do coeficiente de determinação será indicado apenas como forma de ilustrar o quanto o respectivo ajuste foi capaz de explicar os dados.

Tanto o critério de Akaike (*AIC*) quanto o critério bayesiano (*BIC*) aumentam conforme SQE aumenta. Além disso, ambos critérios penalizam modelos com muitas variáveis sendo que valores menores de *AIC* e *BIC* são preferíveis. Como modelos com mais variáveis tendem a produzir menor SQE, mas usam mais parâmetros, a melhor escolha é balancear o ajuste com a quantidade de variáveis.

Outro critério utilizado para indicar a qualidade de um modelo é o log likelihood (*LIK*) que, quanto maior seu valor, melhor o ajuste avaliado. Os resultados dessas estatísticas e do erro padrão das regressões é apresentado na Tabela 29.

Tabela 29: Indicadores de avaliação da qualidade dos modelos de regressão

Indicador	<b>OLS<sub>SB</sub></b>	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
Coeficiente Determinação ( <i>R<sup>2</sup></i> )	0,7958	0,8211	0,8169	0,9239
Log Likelihood ( <i>LIK</i> )	89,255	101,658	97,348	182,005
Critério de Schwarz ( <i>BIC</i> )	– 136,62	– 150,95	– 152,80	– 187,12
Critério de Akaike ( <i>AIC</i> )	– 162,51	– 183,31	– 178,70	– 209,78
Erro Padrão da Regressão ( <i>Se</i> )	0,1538	0,1448	0,1425	0,1303

Fonte: o autor

Os indicadores de desempenho para a regressão linear clássica sem as variáveis indicativas de bairro ( $OLS_{SB}$  - Apêndice 06) foram adicionadas à Tabela 29 para que possam servir de comparação com as outras modelagens.

Todos os modelos tiveram um coeficiente de determinação  $R^2$  altos, sendo a modelagem por regressão geograficamente ponderada a que apresentou o valor maior, indicando conseguir explicar 92,39% da variabilidade dos dados dos imóveis.

Com relação ao log likelihood (*Lik*) os modelos apresentaram crescimento de seu valor, em relação ao modelo  $OLS_{SB}$ , com valor consideravelmente maior para a modelagem GWR, mostrando ser essa, para esse indicador, a melhor escolha.

O modelo de regressão geograficamente ponderada (GWR) foi o que apresentou menor erro padrão no ajuste (*Se*) dentre os modelos analisados.

Quando se observa o critério de informação bayesiano, ou critério de Schwarz (BIC), verifica-se que a regressão geograficamente ponderada (GWR) foi a que apresentou o menor resultado, indicando ter sido a melhor modelagem.

Por fim, o critério de informação de Akaike (AIC) indica que o modelo de regressão geograficamente ponderada, para esses dados, é o melhor ajuste. O valor de AIC para modelagem GWR foi o menor dentre os ajustes realizados.

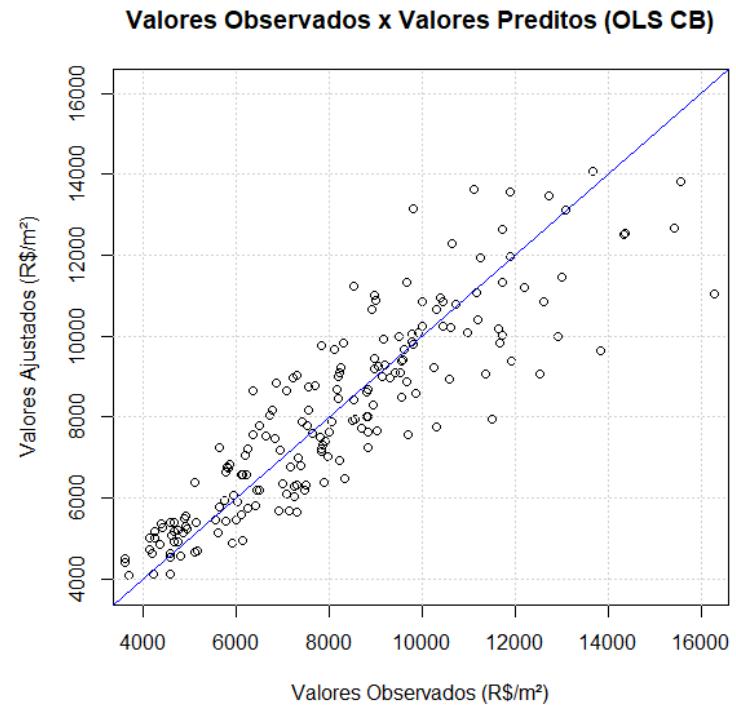
Percebe-se, portanto, que o modelo de regressão geograficamente ponderada ( $GWR_{SB}$ ) foi superior aos modelos de regressão pelos mínimos quadrados ( $OLS_{CB}$ ) e do erro espacial ( $CAR_{SB}$ ) em todos os indicadores de ajuste.

Em geral, quando se realiza avaliação em massa de imóveis, alguns requisitos devem ser cumpridos para que a avaliação seja considerada de boa qualidade. Estes requisitos estão relacionados com o grau de aderência que os valores ajustados pelo modelo e os valores observados no mercado apresentam. Quanto mais próximos estejam os valores ajustados pelo modelo dos valores observados no mercado melhor será a qualidade da avaliação.

Como forma de visualizar graficamente o poder de predição dos modelos, construiu-se o gráfico de dispersão dos valores observados versus valores ajustados para cada uma das modelagens abordadas nesse estudo ( $OLS_{CB}$ ,  $CAR_{SB}$  e  $GWR_{SB}$ ), conforme demonstrado nas Figuras 34, 35 e 36 a seguir.

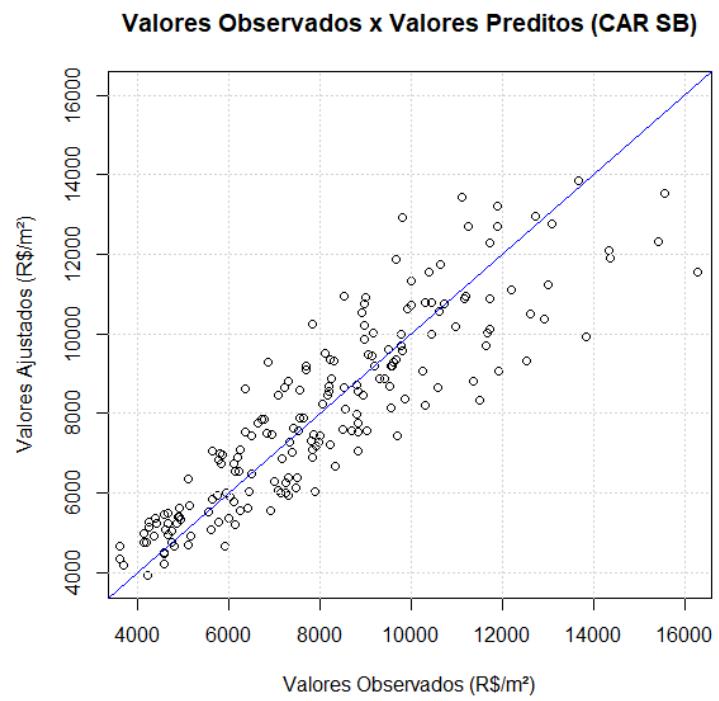
Verifica-se, pelos diagramas que indicam o poder de predição das modelagens, que a regressão geograficamente ponderada (GWR) apresentou-se muito mais ajustada, com pontos menos dispersos, indicando se tratar de um modelo que foi capaz de explicar de forma muito mais eficiente os valores unitários dos imóveis.

Figura 34: Diagrama de dispersão para modelo OLS<sub>CB</sub>



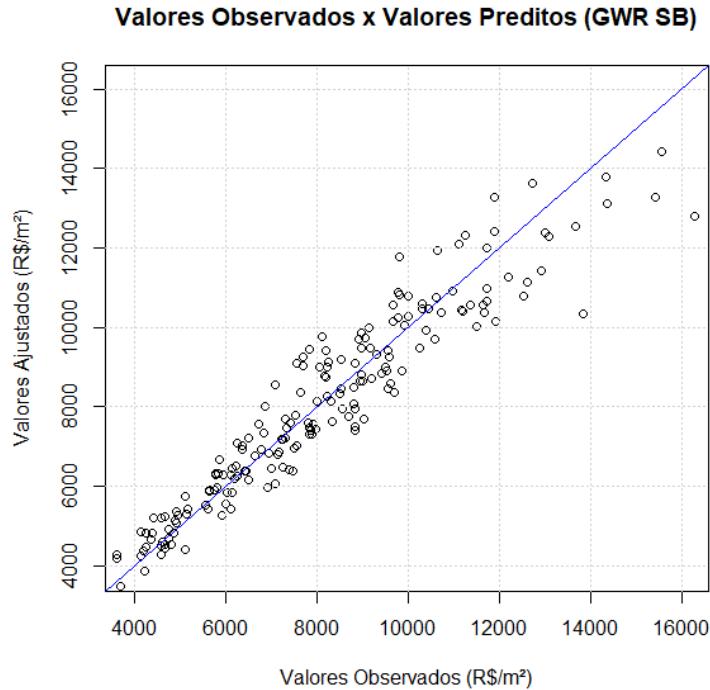
Fonte: o autor

Figura 35: Diagrama de dispersão para modelo CAR<sub>SB</sub>



Fonte: o autor

Figura 36: Diagrama de dispersão para modelo GWR<sub>SB</sub>



Fonte: o autor

Calculou-se, também, para cada modelagem, a raiz do erro quadrático médio, do inglês *root mean square error* (RMSE). Os resultados podem ser vistos na Tabela 30 apresentada. Segundo Marques *et al.* (2012), a raiz quadrada do erro quadrático médio (RMSE) pode ser calculada por meio da equação:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (VU_{OBS} - VU_{AJU})^2}{n-1}} \quad (6.1)$$

Onde VU é o valor unitário observado/ajustado e  $n$  é o número de observações.

Tabela 30: Raiz quadrada do erro quadrático médio da amostra de trabalho

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
RMSE (R\$/m <sup>2</sup> )	1.263,06	1.254,69	848,45

Fonte: o autor

Percebe-se que a modelagem OLS<sub>CB</sub> apresentou o pior desempenho, ao passo que a modelagem GWR<sub>SB</sub>, para esse indicador, apresentou um desempenho muito superior, com valor para o RMSE 33% menor que o valor do modelo OLS<sub>CB</sub>.

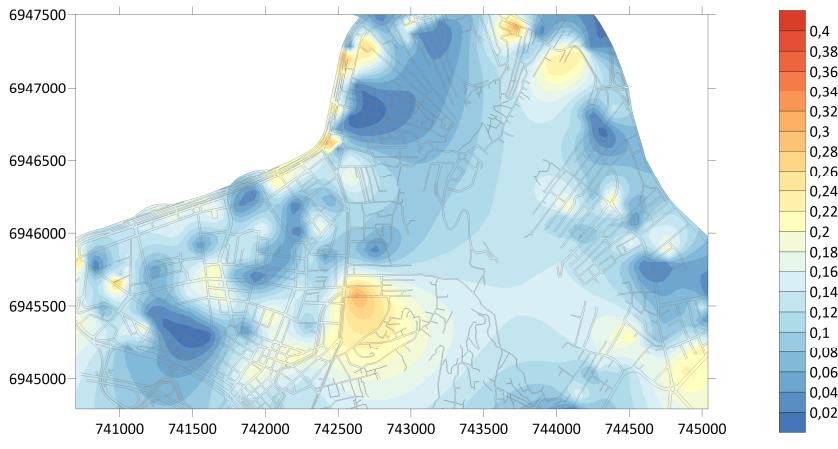
Verificou-se, também, como se deu a dispersão dos erros relativos absolutos dos valores preditos pelas modelagens. Para o cálculo, utilizou-se a equação:

$$ERRO_{REL} = \frac{|VU_{OBS} - VU_{AJU}|}{VU_{OBS}} \quad (6.2)$$

Em que  $VU_{AJU}$  é o valor unitário dos imóveis preditos por cada uma das modelagens e  $VU_{OBS}$  é o valor unitário verdadeiro, observado no mercado.

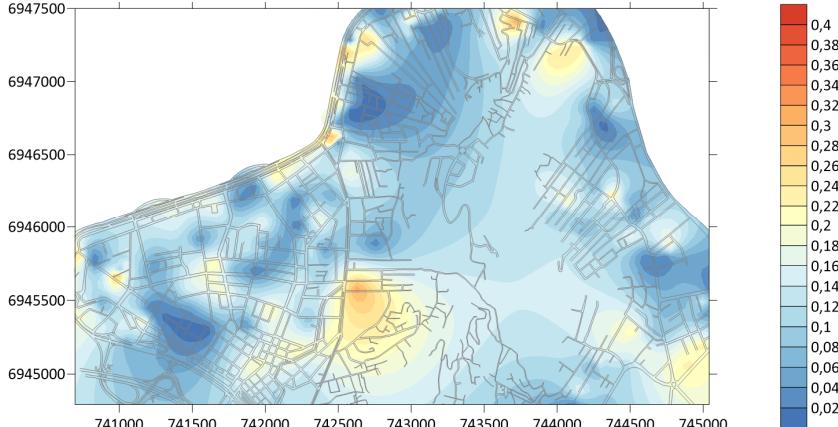
Realizou-se, então, a interpolação linear para geração da superfície de gradientes contendo os erros relativos (absolutos) estimados para cada modelagem:

Figura 37: Superfície de erros relativos absolutos para a modelagem OLS<sub>CB</sub>



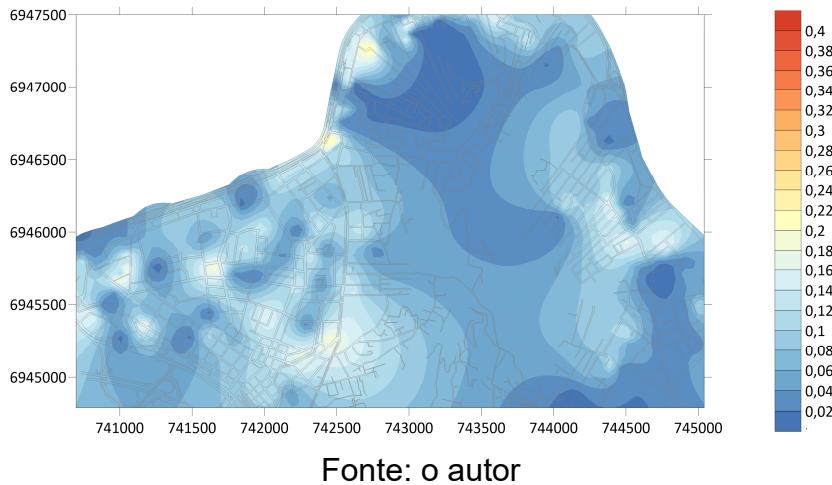
Fonte: o autor

Figura 38: Superfície de erros relativos absolutos para a modelagem CAR<sub>SB</sub>



Fonte: o autor

Figura 39: Superfície de erros relativos absolutos para a modelagem GWR<sub>SB</sub>



Fonte: o autor

Verifica-se, pelas figuras mostradas, que a superfície de erros relativos para a modelagem geograficamente ponderada mostrou-se muito mais uniforme, com erros de menor magnitude, mais próximos de zero, evidenciado pela coloração azul.

Em todas as três modelagens o erro relativo mínimo foi 0,0%. No outro extremo, o erro relativo máximo foi bastante diferente:  $\text{ERRO}_{\text{REL MAX OLS}} = 0,3607$ ;  $\text{ERRO}_{\text{REL MAX CAR}} = 0,3533$  e  $\text{ERRO}_{\text{REL MAX GWR}} = 0,2530$ , evidenciando que a regressão GWR mostrou-se superior as outras duas modelagens analisadas.

Desta forma, pode-se concluir, até esse momento, que a regressão geograficamente ponderada (GWR) possibilitou uma modelagem mais ajustada aos dados dos imóveis da região de estudo, conforme diagramas de predição anteriores, e teve um desempenho superior aos outros dois modelos de regressão em todos os indicadores estatísticos verificados, conforme resultados da Tabela 29.

### 6.6.2 Desempenho dos Modelos de Regressão

Um parâmetro bastante utilizado como medida de *performance* global de uma avaliação em massa é a mediana das razões de avaliação, que é a medida recomendada pela IAAO (*International Association of Assessing Officers*) para monitorar a *performance* global de uma avaliação.

Para se calcular a mediana das razões de avaliação (Med R) foi retirado, aleatoriamente, 35 dados de mercado de imóveis dos três bairros em estudo. Os imóveis da amostra de validação continham 17 dados do bairro Centro, 6 dados do bairro Agronômica e 12 dados do bairro Trindade, conforme visto na Figura 14.

Com a amostra de avaliação em mãos, fez-se a análise dos valores unitários com o objetivo de se verificar a presença de algum dado discrepante que pudesse prejudicar a análise. Contatou-se que o dado AV\_14, cujo valor unitário do imóvel é R\$ 17.500,00, é um *outlier* e, por esse motivo, não será utilizado nos cálculos.

Com os 34 dados de validação restantes (Apêndice 02), fez-se o cálculo da mediana das razões de avaliação (Med R) para os três modelos de regressão.

Importante destacar que os coeficientes utilizados para se fazer a predição dos valores unitários da amostra de validação para os modelos de regressão clássica e regressão espacial são aqueles apresentados nas Tabelas 11 e 20.

Para o cálculo do desempenho da regressão geograficamente ponderada realizou-se, diretamente no software GWR4, a predição dos valores unitários da amostra de validação. O software GWR4 fornece, para cada uma das amostras de validação, tendo como base o *kernel* e *bandwidth* selecionados, uma equação de predição, com os respectivos valores da estatística *t*, do erro padrão e do *R*<sup>2</sup> local.

Como parâmetro de dispersão horizontal da avaliação, a medida utilizada no estudo para avaliar a uniformidade da avaliação é o coeficiente de dispersão (COD), considerando o caso de imóveis urbanos (apartamentos) em regiões heterogêneas.

Por fim, para medir a equidade vertical da avaliação, utilizou-se o parâmetro chamado de diferencial relativo de preço ou PRD, que busca detectar diferenças sistemáticas na forma como são avaliados os imóveis de alto e de baixo valor.

Conforme recomendações da IAAO (*International Association of Assessing Officers*), para os três parâmetros, os limites estão definidos na Tabela 31.

Tabela 31: Limites recomendados para os parâmetros de avaliação em massa

Indicador	Recomendação
Mediana das Razões de Avaliação (Med R)	$0,90 < \text{Med R} < 1,10$
Coeficiente de Dispersão (COD)	$\text{COD} \leq 15\%$
Diferencial Relativo ao Preço (PRD)	$0,98 \leq \text{PRD} \leq 1,03$

Fonte: IAAO (2013)

Nesse caso, se está avaliando o coeficiente de dispersão da mediana das razões de avaliação (COD) para imóveis em zonas heterogêneas. Os resultados dos cálculos dos indicadores apresentados na Tabela 31, para os três modelos de regressão analisados (OLS<sub>CB</sub>, CAR<sub>SB</sub> e GWR<sub>SB</sub>), são apresentados na Tabela 32.

Tabela 32: Valores calculados para os indicadores de avaliação em massa

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
Med R	0,949	0,968	0,957
COD (%)	12,25%	12,22%	10,57%
PRD	1,018	1,017	1,015

Fonte: o autor

Com relação à mediana das razões de avaliação (Med R), todos os modelos tiveram seus parâmetros dentro dos limites recomendados pela IAAO, estando o modelo CAR<sub>SB</sub> com o resultado mais próximo de 1,0, indicando que os valores preditos por esse modelo possuem razões de avaliação igualmente distribuídos, com quantidades praticamente iguais de dados ajustados para mais e para menos em relação ao valor unitário dos imóveis que compõe a amostra da validação.

Com relação à medida de dispersão horizontal (COD), todas as três modelagens tiveram seus valores inferiores ao máximo recomendado pela IAAO, entretanto a regressão geograficamente ponderada (GWR) foi a que apresentou maior uniformidade da avaliação dos imóveis em toda a área de estudo.

Por fim, com relação ao cálculo da equidade vertical (PRD), todas as modelagens apresentaram valores dentro dos limites recomendados pela IAAO, indicando que não houve progressividade e nem regressividade em nenhum dos três modelos, contudo, a regressão GWR teve desempenho ligeiramente superior nesse indicador.

Deve-se destacar que o valor dos indicadores para a regressão geograficamente ponderada que considerava em seu cálculo a largura de banda de 278,42 m teve o melhor desempenho, em relação às outras modelagens, sendo capaz de capturar de forma muito mais eficiente a influência dos vizinhos, comprovando a importância destes na predição do valor de mercado de um imóvel.

Calculou-se, também, a raiz quadrada do erro quadrático médio dos valores observados e preditos pela amostra de validação para cada uma das modelagens (OLS<sub>CB</sub>, CAR<sub>SB</sub>, GWR<sub>SB</sub>). Os resultados podem ser vistos na Tabela 33.

Tabela 33: Raiz quadrada do erro quadrático médio da amostra de validação

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
RMSE (R\$/m <sup>2</sup> )	1.433,86	1.412,27	1.321,25

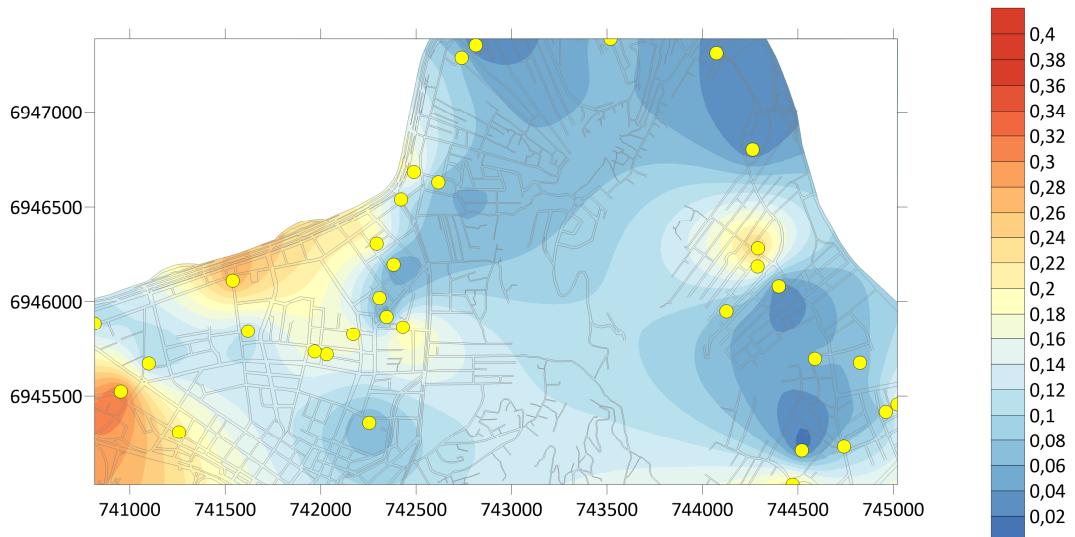
Fonte: o autor

Observa-se semelhança no valor da raiz do erro quadrático médio (RMSE) para as modelagens OLS<sub>CB</sub> e CAR<sub>SB</sub>, decrescendo destas para a modelagem GWR<sub>SB</sub>. Nesse caso, a modelagem que apresentou o menor valor de RMSE foi a GWR<sub>SB</sub>. Esse fenômeno novamente corrobora com a conclusão de que o valor de um determinado imóvel é fortemente afetado pelo valor dos imóveis do seu entorno.

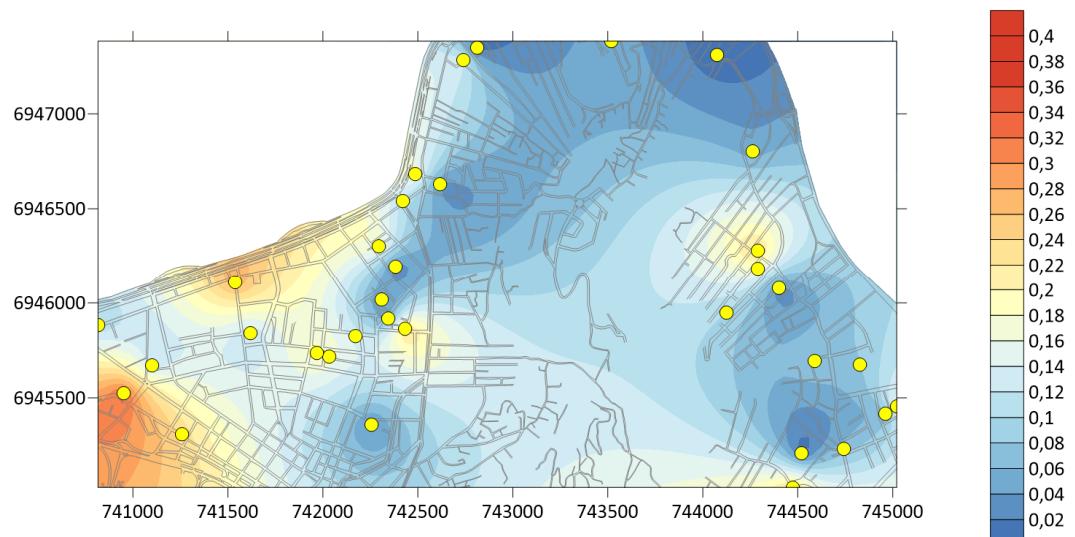
Gerou-se, também, para cada uma das modelagens da Tabela 33, uma superfície contendo os erros relativos (absolutos) dos valores unitários observados e dos valores unitários preditos pelos modelos, conforme Figuras 40 A a D:

Figura 40 A a D: Erros relativos da amostra de validação para as modelagens

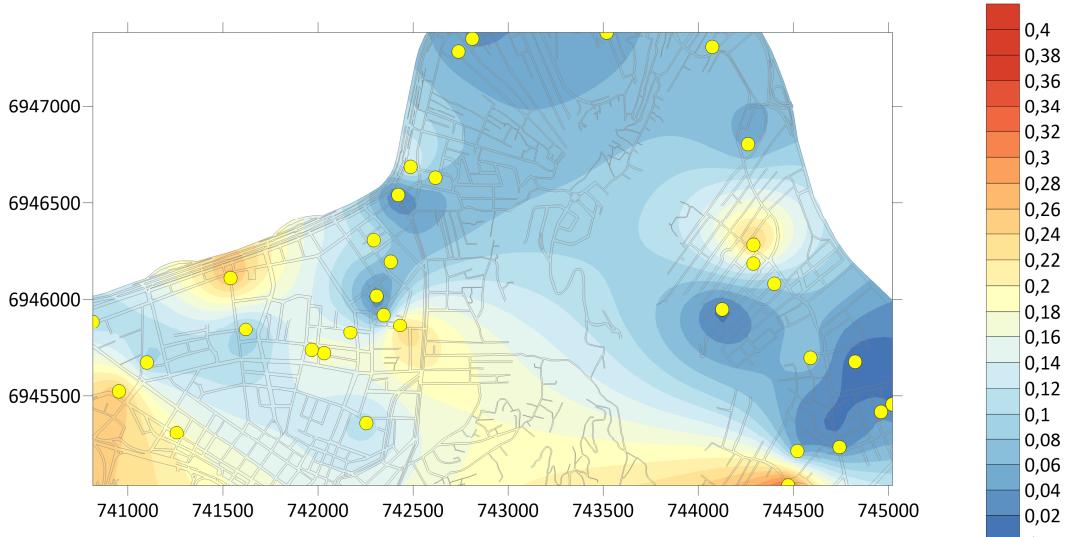
A) Modelagem OLS<sub>CB</sub>



B) Modelagem CAR<sub>SB</sub>



### C) Modelagem GWR<sub>SB</sub>



Fonte: o autor

Na Tabela 34 é possível verificar, para cada modelagem, o erro relativo absoluto médio obtido para as 34 amostras de validação consideradas nesse estudo.

Tabela 34: Valor do erro relativo absoluto médio das amostras de validação

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
<b>ERRO<sub>REL</sub></b>	0,120	0,118	0,107

Fonte: o autor

Observa-se uma pequena redução do erro relativo médio quando se compara a regressão clássica com a regressão do erro espacial e uma redução desta para a regressão geograficamente ponderada. Essa situação é perceptível quando se analisa os mapas de gradientes, nos quais se observa similaridade entre as regressões OLS<sub>CB</sub> e CAR<sub>SB</sub>. No entanto, se vê uma redução grande dos erros relativos quando se adota a regressão GWR<sub>SB</sub>, evidenciado pelo aumento das regiões em azul e redução das regiões vermelhas nas superfícies da Figura 40.

Os resultados desse tópico indicam que, embora todos os modelos tenham apresentado parâmetros dentro das recomendações da IAAO, a modelagem GWR mostrou-se superior às outras duas modelagens em relação ao seu desempenho por ter apresentado o menor valor para o diferencial relativo ao preço (PRD) e um coeficiente de dispersão horizontal (COD) menor que as outras duas modelagens.

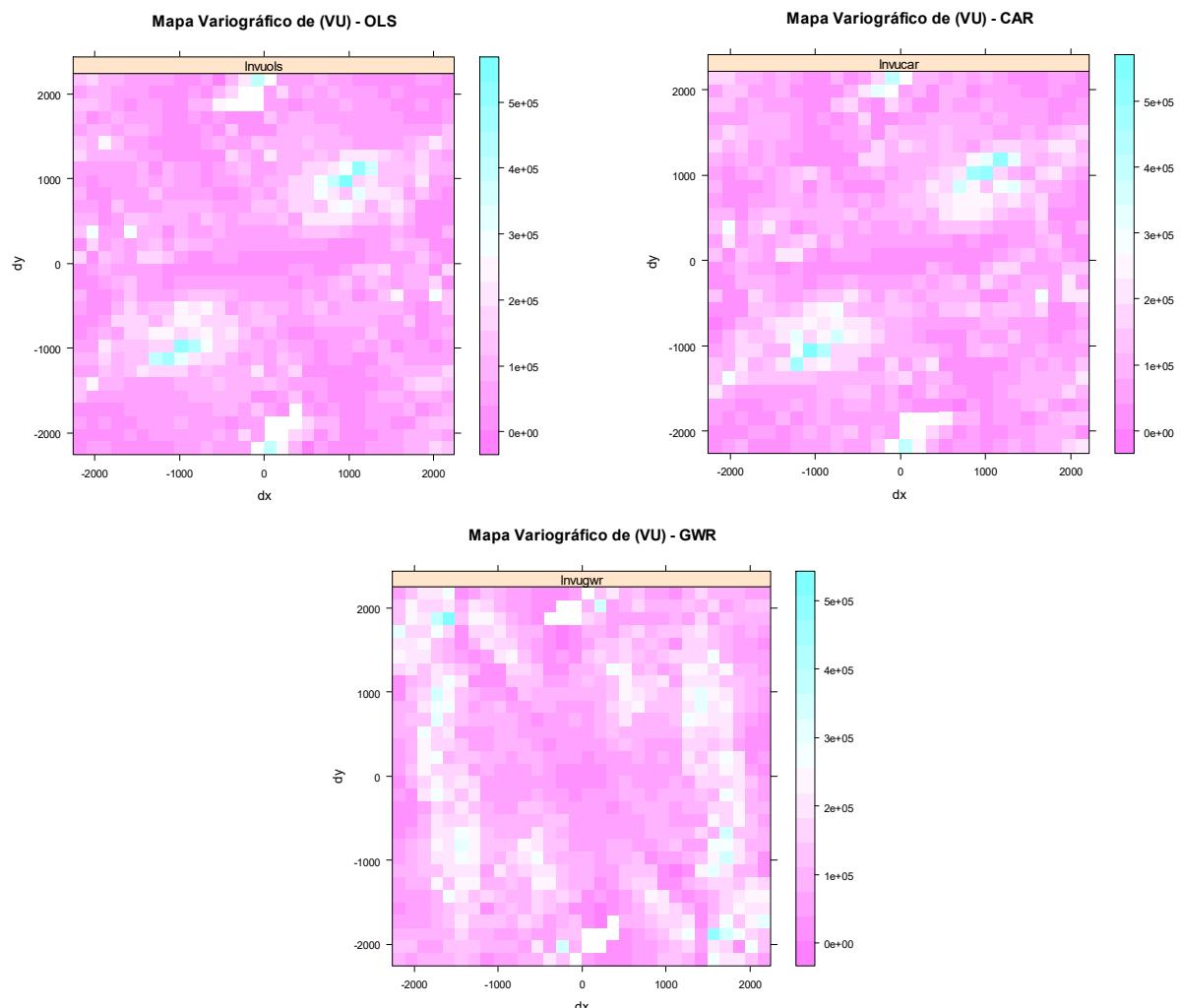
## 6.7 SUPERFÍCIES DE VALORES UNITÁRIOS: KRIGAGEM ORDINÁRIA

O mapa variográfico é utilizado para observar todas as direções do espaço em um mesmo gráfico, a fim de auxiliar o pesquisador na determinação da direção principal do fenômeno espacial, ou seja, se há ou não anisotropia na área.

Posteriormente, ao ser detectado se houve anisotropia ou não, o variograma experimental é calculado ao longo das direções de anisotropia reconhecidas. Tanto os mapas variográficos quanto os variogramas foram obtidos por meio do *software R* na sua versão 3.5.3. e os códigos são apresentados no Apêndice 05 desse trabalho.

Quando um fenômeno em estudo é isotrópico, ou seja, apresenta o mesmo semivariograma em todas as direções, um único semivariograma teórico pode ser representativo da estrutura de variação espacial da variável avaliada.

Figura 41 A a C: Mapas variográficos dos efeitos direcionais dos modelos



Fonte: o autor

Observando os mapas variográficos dos três modelos de regressão gerados, apresentados na Figura 41, constata-se que não existem tendências direcionais quando se analisa o comportamento dos pixels para as diversas direções do mapa. Dessa forma, o fenômeno espacial em estudo é considerado isotrópico.

Nesse caso, o semivariograma teórico representativo da estrutura de variação espacial da variável avaliada, utilizado ao longo desse estudo, foi o semivariograma exponencial, por apresentar o menor erro quadrático médio (MSE) e maior dependência espacial, como poderá ser visto nos tópicos seguintes.

### 6.7.1 Considerações Iniciais: Superfícies de Valores

Quando se deseja produzir uma superfície de valores de imóveis a partir da interpolação por krigagem ordinária, deve-se, inicialmente, fazer a homogeneização dos valores preditos pelos modelos de regressão, de modo a obter uma superfície contínua de predição dos valores unitários dos imóveis da área em estudo. Essas superfícies foram construídas no software Surfer na sua versão 15.

Para isso, definiu-se o imóvel paradigma representativo da área em estudo. O critério para definição do apartamento paradigma foi buscar a média aproximada da área privativa dos imóveis da amostra e a moda para as demais características.

Tabela 35: Apartamento paradigma para homegeneização

	<b>AP</b>	<b>ND</b>	<b>NB</b>	<b>NG</b>	<b>PS</b>	<b>PC</b>
Paradigma	110	3	2	1	0	2

Fonte: o autor

Realizada a homogeneização e o cálculo dos novos valores unitários para os imóveis, procedeu-se com a definição dos semivariogramas que darão suporte para criação das superfícies de valores unitários para as três modelagens em estudo.

Importante mencionar que a homogeneização dos valores unitários para os modelos de regressão clássica e regressão espacial foi realizada, para o imóvel paradigma acima, diretamente sobre o modelo global das Tabelas 11 e 20.

A homogeneização dos valores unitários para a regressão geograficamente ponderada para cada ponto amostral ocorreu sobre a regressão do próprio ponto, substituindo-se o imóvel paradigma do bairro em que se encontrava a amostra na sua respectiva regressão, variando apenas a distância ao polo valorizador.

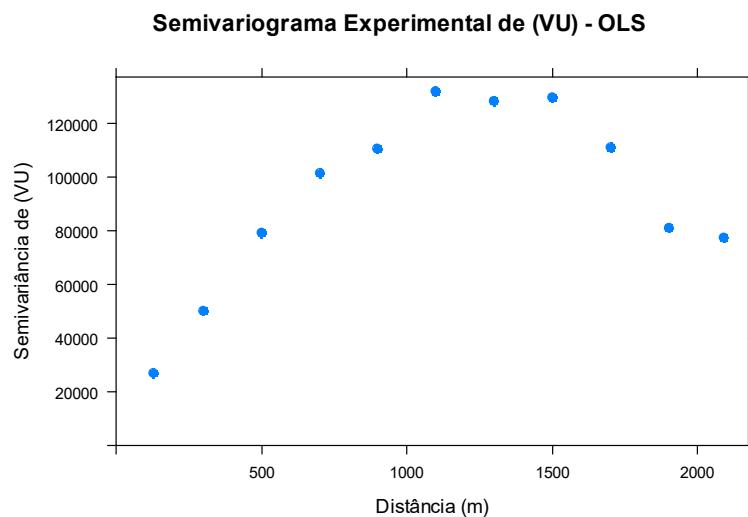
## 6.7.2 Superfície de Predição da Regressão Linear Clássica

Definido o imóvel paradigma que representa os bairros da área em estudo, construiu-se o semivariograma omnidirecional experimental. Esse semivariograma experimental permitiu extrair informações para construção dos semivariogramas teóricos que forneceram os valores de efeito pepita, patamar e alcance para, enfim, se modelar por krigagem ordinária as superfícies de valores homogeneizados.

### 6.7.2.1 Modelagem dos Semivariogramas OLS<sub>CB</sub>

Nessa etapa, o semivariograma experimental foi construído sobre os valores unitários homogeneizados preditos pelo modelo de regressão por mínimos quadrados ordinários (OLS<sub>CB</sub>). Esse semivariograma permitiu extrair informações relacionadas ao efeito pepita, contribuição, patamar e alcance, que serão utilizadas para modelagem do semivariograma teórico, como pode ser visto na Figura 42.

Figura 42: Semivariograma experimental dos valores unitários



Fonte: o autor

Os modelos utilizados para construção dos semivariogramas teóricos foram gaussiano, esférico e exponencial e estão disponíveis no Apêndice 07.

Segundo Guimarães (2004), a dependência espacial pode ser avaliada pela razão entre o efeito pepita ( $C_0$ ) e o patamar (C) e fica caracterizada uma forte dependência espacial para valores inferiores a 25%, moderada dependência

espacial para valores entre 25% e 75% e fraca dependência espacial para valores acima de 75%. A Tabela 36 mostra esses parâmetros para cada modelo teórico.

Tabela 36: Parâmetros dos semivariogramas teóricos do ajuste OLS<sub>CB</sub>

Parâmetros	Esférico	Exponencial	Gaussiano
Efeito Pepita ( $C_0$ )	8.749	5.788	18.603
Contribuição ( $C'$ )	118.785	161.700	101.199
Patamar (C)	127.534	167.488	119.802
Coeficiente ( $C_0/C$ )	6,86%	3,45%	15,52%
Alcance (A)	1.236,96	874,81	504,47
Dependência	Forte	Forte	Forte

Fonte: o autor

Observa-se que o maior alcance ocorreu com o modelo esférico e o menor ocorreu com o modelo gaussiano. Verifica-se, ainda, que a menor razão  $C_0/C$  ocorreu no modelo exponencial, o que indica que, nesse modelo teórico, 3,45% da variabilidade total dos resíduos é devido ao componente aleatório e que 96,55% são explicados pela presença de autocorrelação espacial nos resíduos.

Como forma de auxiliar na escolha do melhor modelo teórico ajustado, realizou-se a validação cruzada com auxílio do software R. A validação cruzada é realizada com base nos valores unitários (VU) preditos pelo modelo de regressão pelos mínimos quadrados ordinários (OLS<sub>CB</sub>) e pelos modelos teóricos considerados.

Calculou-se, para cada um dos modelos teóricos, o valor do coeficiente de determinação ( $R^2$ ), o erro quadrático médio, do inglês *mean square error* (MSE) e a raiz quadrada do erro quadrático médio, do inglês *root mean square error* (RMSE). Nesse caso, o melhor modelo é aquele que apresentar os menores erros e o maior  $R^2$ . Os resultados podem ser vistos na Tabela 37.

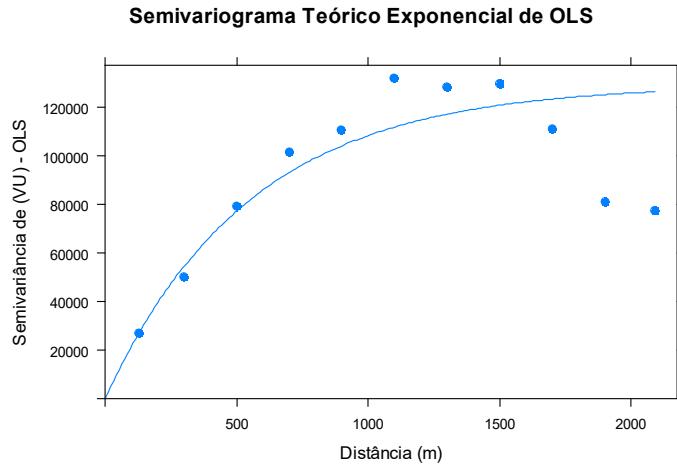
Tabela 37: Validação cruzada entre os modelos teóricos e o ajuste OLS<sub>CB</sub>

Parâmetros	Esférico	Exponencial	Gaussiano
$R^2$	0,9775	0,9782	0,9748
MSE	10.569,90	10.237,39	11.804,82
RMSE	102,81	101,18	108,65

Fonte: o autor

Como o modelo exponencial apresentou o menor erro quadrático médio (MSE), o maior coeficiente de determinação ( $R^2$ ) e a maior dependência espacial, optou-se por utilizá-lo na etapa seguinte. O semivariograma é visto na Figura 43:

Figura 43: Semivariograma teórico exponencial para regressão OLS<sub>CB</sub>



Fonte: o autor

#### 6.7.2.2 Superfície de Valores Preditos pelo Modelo OLS<sub>CB</sub>

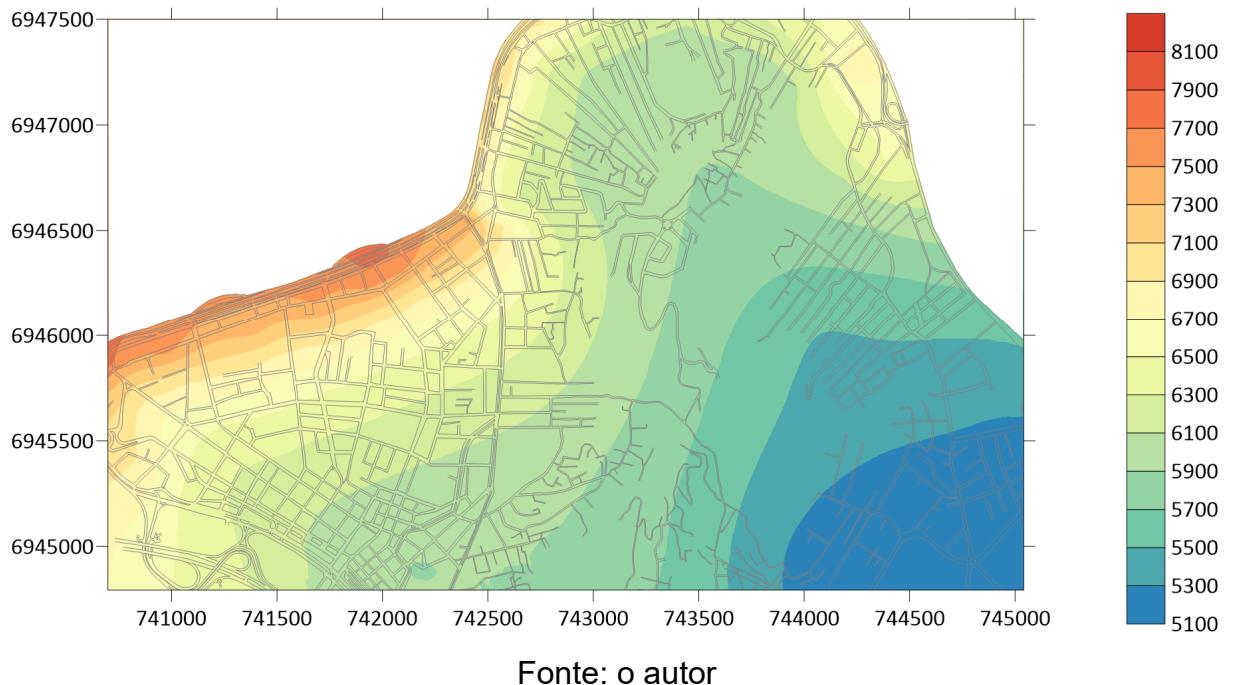
Para construção das superfícies de valores ajustados pelo modelo utilizou-se os parâmetros obtidos no semivariograma do modelo teórico exponencial por apresentar o menor erro quadrático médio e possuir a maior dependência espacial.

Construiu-se, para a região de estudo, uma superfície de predição dos valores unitários homogeneizados e uma superfície de predição do erro da krigagem ordinária. Na sequência é apresentada a superfície de predição dos valores unitários com base no imóvel paradigma, em R\$/m<sup>2</sup>, obtidos por krigagem ordinária.

Na Figura 44 é possível verificar mapa da variável (VU) espacializada a partir da krigagem. Observa-se, pela superfície de predição dos valores unitários homogeneizados, que nas proximidades da Avenida Beira Mar Norte estão localizados os imóveis mais valorizados da região em estudo. Verifica-se que, para a Agronômica, a modelagem não se mostrou tão eficiente (cores mais claras).

Observa-se, pela imagem, um degrau entre os bairros Centro e Agronômica. Ele acontece por causa das variáveis dicotômicas: no Centro há uma valorização em relação aos outros dois bairros. Este degrau não existe entre a Agronômica e a Trindade porque a desvalorização deles, em relação ao Centro, é muito semelhante.

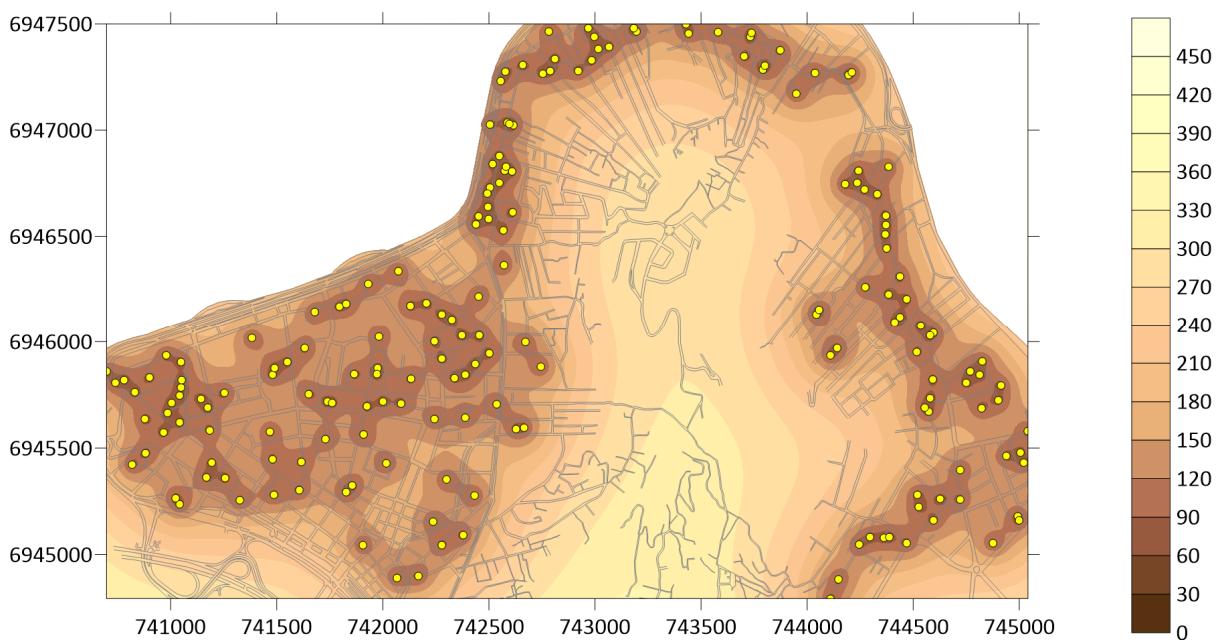
Figura 44: Superfície de predição dos valores unitários para OLS<sub>CB</sub>



Fonte: o autor

Já na Figura 45, tem-se a superfície de predição do erro da krigagem dos valores unitários preditos pela modelagem clássica. Com relação à superfície de predição do erro, constata-se que no entorno dos pontos da amostra o erro da estimativa é menor e vai aumentando à medida que se afasta desses pontos.

Figura 45: Superfície predição do erro para OLS<sub>CB</sub>



Fonte: o autor

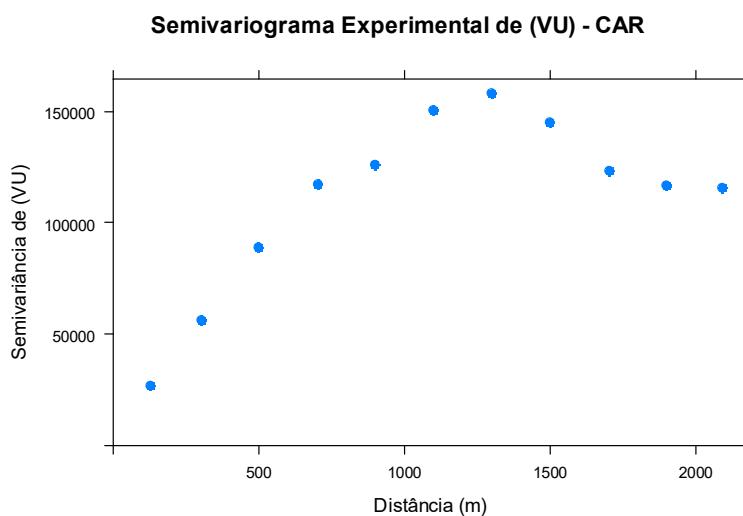
### 6.7.3 Superfície de Predição da Regressão do Erro Espacial

Nessa etapa buscou-se modelar superfícies de valores a partir dos resultados obtidos na modelagem por regressão do erro espacial ( $CAR_{SB}$ ) para a região dos bairros em estudo. Fez-se, inicialmente, a modelagem do semivariograma experimental dos valores unitários (VU) ajustados pelo modelo de regressão CAR.

#### 6.7.3.1 Modelagem dos Semivariogramas $CAR_{SB}$

Gerou-se o semivariograma omnidirecional experimental dos valores unitários homogeneizados pelo modelo de regressão do erro espacial ( $CAR_{SB}$ ). O semivariograma experimental construído na Figura 46 permitiu extrair informações que serão utilizadas para modelagem do semivariograma teórico.

Figura 46: Semivariograma experimental dos valores unitários



Fonte: o autor

Os modelos utilizados para construção dos semivariogramas teóricos foram gaussiano, esférico e exponencial e estão disponíveis no Apêndice 07.

Novamente, a dependência espacial foi avaliada pela razão entre o efeito pepita ( $C_0$ ) e o patamar ( $C$ ). Essa relação permite que se verifique, percentualmente, quanto da variabilidade total é explicada pela presença de autocorrelação espacial dos resíduos, fazendo  $(1 - C_0/C)$ . A Tabela 38 mostra os parâmetros calculados.

Tabela 38: Parâmetros dos semivariogramas teóricos do ajuste CAR<sub>SB</sub>

Parâmetros	Esférico	Exponencial	Gaussiano
Efeito Pepita ( $C_0$ )	1.444	0.000	13.100
Contribuição ( $C'$ )	144.153	192.315	119.183
Patamar (C)	145.597	192.315	132.283
Coeficiente ( $C_0/C$ )	1,00%	0,00%	9,90%
Alcance (A)	1.159,47	835,94	449,73
Dependência	Forte	Forte	Forte

Fonte: o autor

Observa-se que o maior alcance ocorreu com o modelo esférico e o menor ocorreu com o modelo gaussiano. Verifica-se, ainda, que o menor efeito pepita ocorreu no modelo exponencial, que apresentou forte dependência espacial.

Os resultados indicam que, no modelo exponencial, 100% da variabilidade total dos resíduos é explicada pela presença de autocorrelação espacial.

Como forma de auxiliar na escolha do melhor modelo teórico ajustado, realizou-se a validação cruzada com auxílio do software R. A validação cruzada é realizada com base nos valores unitários (VU) preditos pelo modelo de regressão do erro espacial (CAR<sub>SB</sub>) e pelos modelos teóricos considerados nesse estudo.

Calculou-se, para cada um dos modelos teóricos, o valor do coeficiente de determinação ( $R^2$ ), o erro quadrático médio, do inglês *mean square error* (MSE) e a raiz quadrada do erro quadrático médio, do inglês *root mean square error* (RMSE).

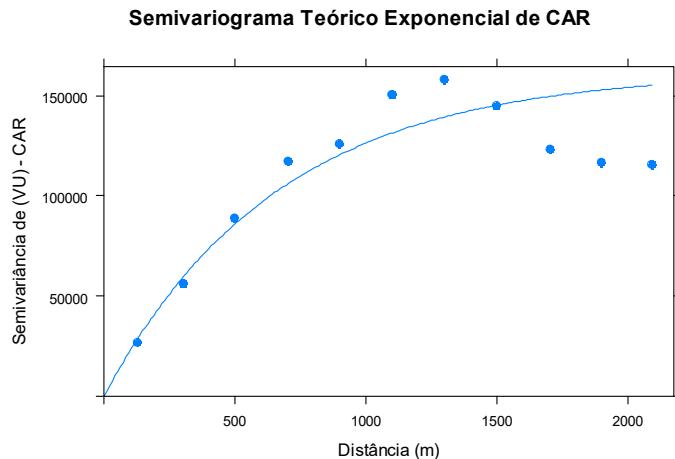
Tabela 39: Validação cruzada entre os modelos teóricos e o ajuste CAR<sub>SB</sub>

Parâmetros	Esférico	Exponencial	Gaussiano
$R^2$	0,9858	0,9866	0,9813
MSE	6.661,82	6.302,78	8.656,44
RMSE	81,62	79,39	93,04

Fonte: o autor

O modelo exponencial apresentou o menor erro quadrático médio (MSE) e a mais forte dependência espacial. Optou-se por utilizar esse modelo na etapa seguinte. O semivariograma teórico exponencial pode ser visto na Figura 47.

Figura 47: Semivariograma teórico exponencial para regressão CAR<sub>SB</sub>



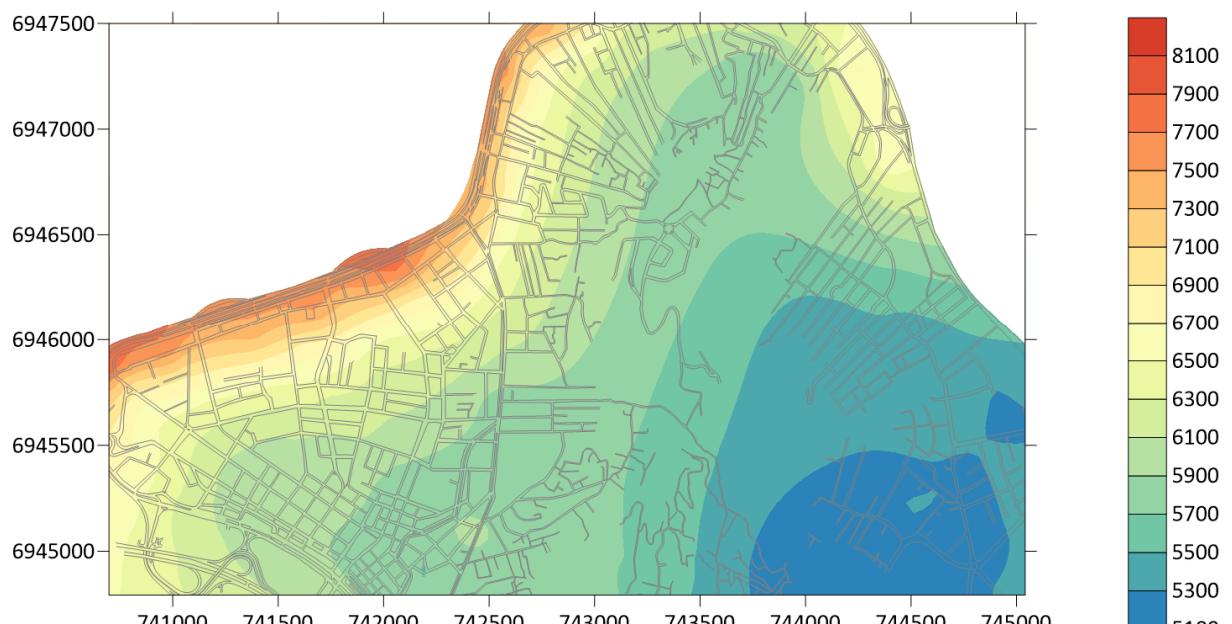
Fonte: o autor

#### 6.7.3.2 Superfície de Valores Preditos pelo Modelo CAR<sub>SB</sub>

Para construção das superfícies de valores ajustados pelo modelo utilizou-se os parâmetros obtidos no semivariograma do modelo teórico exponencial por apresentar o menor erro quadrático médio e possuir a maior dependência espacial.

Na sequência é apresentada a superfície de predição dos valores unitários com base no imóvel paradigma, em R\$/m<sup>2</sup>, obtidos por krigagem ordinária. Essa modelagem se mostrou mais adequada e não apresentou degrau, como a anterior.

Figura 48: Superfície de predição dos valores unitários para CAR<sub>SB</sub>

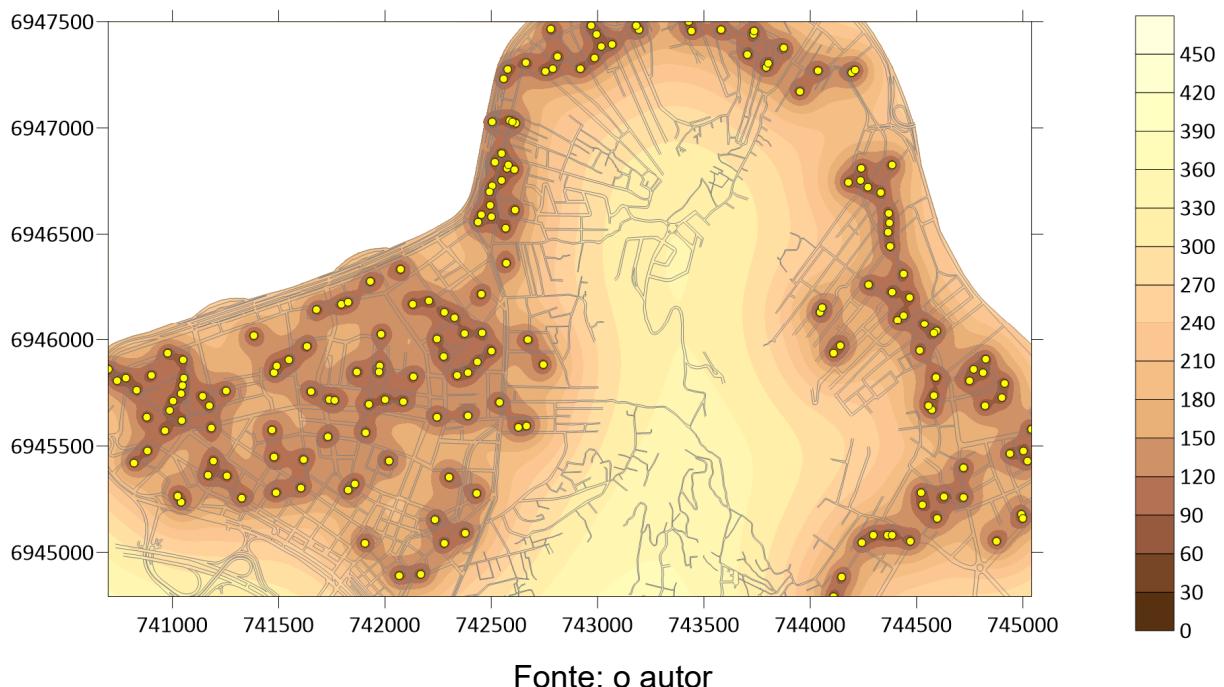


Fonte: o autor

Pode-se perceber, pela análise da superfície de predição dos valores unitários (Figura 48), que a Avenida Beira Mar tem forte influência sobre o valor unitário dos imóveis na região de Florianópolis. Verifica-se, ainda, que a superfície gerada para o bairro Agronômica teve sua representação mais próxima do esperado, com imóveis mais valorizados, quando comparada com a modelagem OLS<sub>CB</sub>.

Com relação à superfície de predição do erro (Figura 49), constata-se, novamente, que, próximo aos pontos da amostra coletada, o erro da estimativa é menor e vai aumentando à medida que se afasta desses pontos.

Figura 49: Superfície predição do erro para CAR<sub>SB</sub>



Fonte: o autor

Esse fenômeno demonstra a importância de se ter um número de dados em quantidade suficiente para oferecer uma espacialização que cubra boa parte da área em estudo a fim de se obter superfícies de valores unitários confiáveis e precisas.

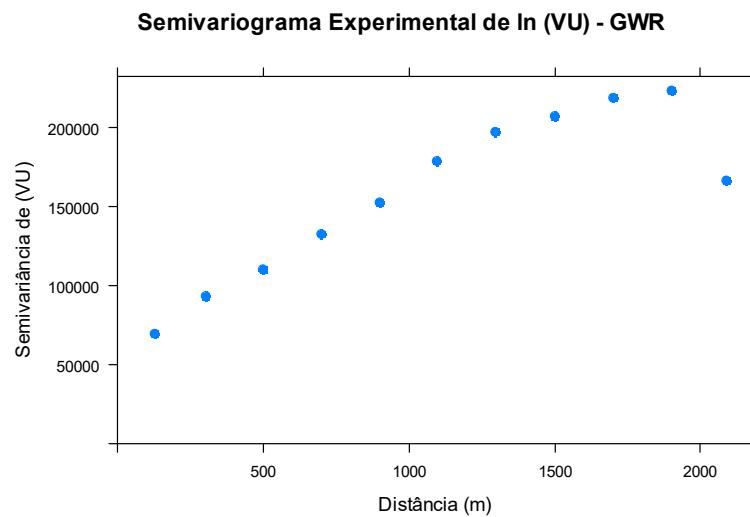
#### 6.7.4 Superfície de Predição da Regressão Geograficamente Ponderada

Nessa etapa buscou-se modelar superfícies de valores a partir dos resultados obtidos na modelagem por regressão geograficamente ponderada (GWR<sub>SB</sub>). Fez-se, inicialmente, a modelagem do semivariograma experimental dos valores unitários homogeneizados pelo modelo de regressão pelo método GWR.

#### 6.7.4.1 Modelagem dos Semivariogramas GWR<sub>SB</sub>

Gerou-se o semivariograma experimental dos valores unitários ajustados pelo modelo de regressão geograficamente ponderada (GWR<sub>SB</sub>). O semivariograma experimental construído permitiu extrair informações importantes que serão utilizadas nas etapas seguintes para a definição do semivariograma teórico.

Figura 50: Semivariograma experimental dos valores unitários



Fonte: o autor

Foi avaliada a dependência espacial pela razão entre o efeito pepita ( $C_0$ ) e o patamar (C) e é classificada como forte dependência espacial para valores inferiores a 25%, moderada dependência espacial para valores entre 25% e 75% e fraca dependência espacial para valores acima de 75% (GUIMARÃES, 2004).

Tabela 40: Parâmetros dos semivariogramas teóricos do ajuste GWR<sub>SB</sub>

Parâmetros	Esférico	Exponencial	Gaussiano
Efeito Pepita ( $C_0$ )	63.879	40.681	72.029
Contribuição (C')	296.847	350.832	124.422
Patamar (C)	360.726	391.513	196.451
Coeficiente ( $C_0/C$ )	17,70%	10,40%	36,60%
Alcance (A)	4.585,52	3.008,65	813,53
Dependência	Forte	Forte	Moderada

Fonte: o autor

Observa-se que o maior alcance ocorreu com o modelo esférico e o menor ocorreu com o modelo gaussiano. Verifica-se, ainda, que o modelo teórico exponencial apresentou a mais forte dependência espacial em que 89,60% da variabilidade é explicada pela presença de autocorrelação espacial nos resíduos.

Como forma de auxiliar na escolha do melhor modelo teórico ajustado, realizou-se a validação cruzada com auxílio do software R.

Calculou-se, para cada um dos modelos teóricos, o valor do coeficiente de determinação ( $R^2$ ), o erro quadrático médio (MSE) e a raiz quadrada do erro quadrático médio, (RMSE). Os resultados podem ser vistos pela Tabela 41.

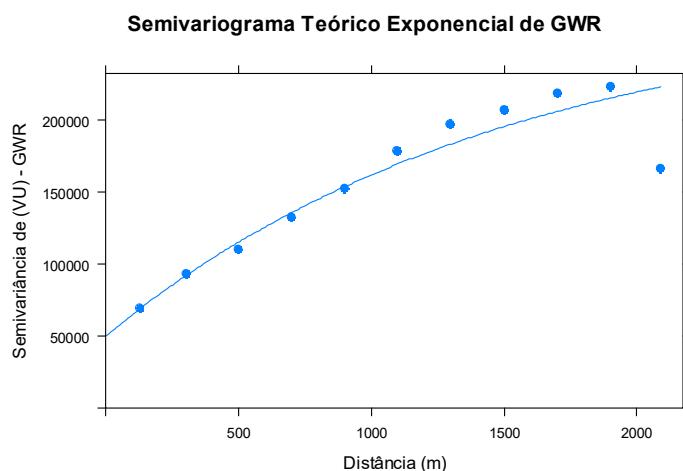
Tabela 41: Validação cruzada entre os modelos teóricos e o ajuste GWR<sub>SB</sub>

<b>Parâmetros</b>	<b>Esférico</b>	<b>Exponencial</b>	<b>Gaussiano</b>
$R^2$	0,9236	0,9264	0,8978
MSE	40.610,31	39.239,64	53.657,08
RMSE	201,52	198,09	231,64

Fonte: o autor

O modelo teórico exponencial foi utilizado na etapa seguinte por apresentar o menor erro quadrático médio (MSE), o maior coeficiente de determinação e forte dependência espacial. O semivariograma exponencial, calculado sobre os valores unitários homogeneizados pela regressão GWR<sub>SB</sub>, pode ser visto na Figura 51.

Figura 51: Semivariograma teórico exponencial para regressão GWR<sub>SB</sub>



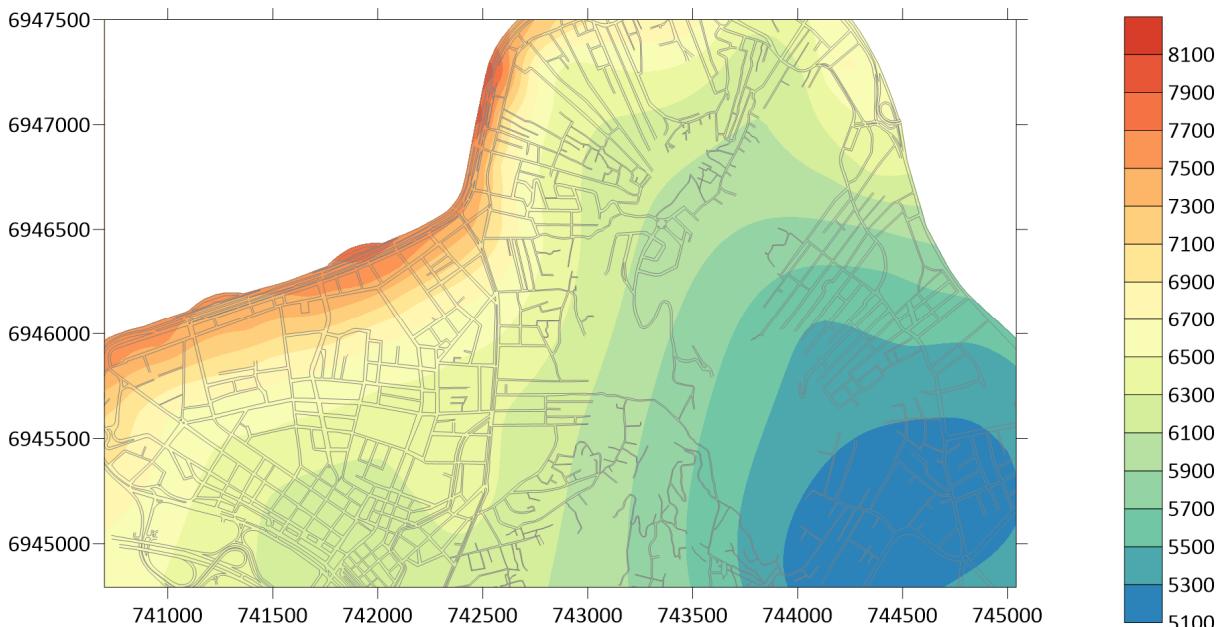
Fonte: o autor

#### 6.7.4.2 Superfície de Valores Preditos pelo Modelo GWR<sub>SB</sub>

Para construção das superfícies de valores ajustados pelo modelo utilizou-se os parâmetros obtidos no semivariograma do modelo teórico exponencial por apresentar a maior dependência espacial e o menor erro quadrático médio.

Construiu-se, para a região de estudo, uma superfície de predição dos valores unitários homogeneizados e uma superfície de predição do erro da krigagem ordinária. Na sequência é apresentada a superfície de predição dos valores unitários com base no imóvel paradigma, em R\$/m<sup>2</sup>, obtidos por krigagem ordinária.

Figura 52: Superfície de predição dos valores unitários para GWR<sub>SB</sub>



Fonte: o autor

Pode-se constatar, pelas regiões mais quentes espalhadas por boa parte da extensão do Centro, que a superfície GWR<sub>SB</sub> foi capaz de explicar de maneira mais próxima da realidade o valor unitário dos imóveis deste bairro, tendo em vista que, comumente, o Centro possui os apartamentos mais valorizados de Florianópolis.

O mesmo ocorre com a krigagem do bairro Agronômica, onde a faixa da avenida Beira Mar concentra apartamentos novos e valorizados em Florianópolis.

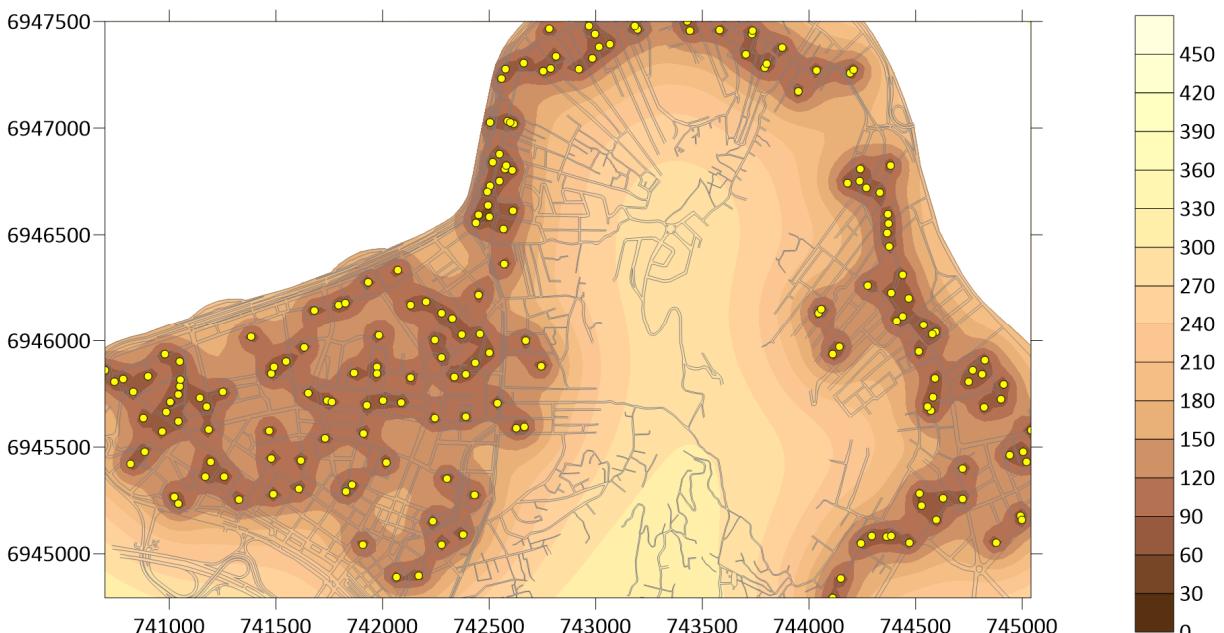
Nota-se, ainda, que as superfícies geradas pelos modelos CAR<sub>SB</sub> e GWR<sub>SB</sub>, por não incluírem variáveis dicótomicas indicativas de bairro, como em OLS<sub>CB</sub>, não apresentam degraus, como os vistos na modelagem da Figura 44. Observa-se, também, pela amplitude dos valores unitários preditos pela superfície GWR<sub>SB</sub>, que a

krigagem ordinárias conseguiu representar a variação dos valores unitários dos imóveis que ocorre na área em estudo de forma mais detalhada.

Pode-se perceber que, nas duas modelagens anteriores ( $OLS_{CB}$  e  $CAR_{SB}$ ), os valores unitários homogeneizados para o imóvel paradigma considerado tiveram uma variação entre R\$ 5.100,00 e R\$ 7.800,00 por metro quadrado, ao passo que, na regressão geograficamente ponderada ( $GWR_{SB}$ ) os valores unitários variaram entre R\$ 5.000,00 e R\$ 8.100,00 reais por metro quadrado, aproximadamente.

Verifica-se, também, que a superfície de predição do erro da krigagem para o modelo  $GWR_{SB}$  apresentou menores valores de erro no entorno das observações, aumentando à medida que ocorre o afastamento desses pontos. Essa modelagem apresentou de forma mais intensa as regiões escuras, como visto na Figura 53.

Figura 53: Superfície predição do erro para  $GWR_{SB}$

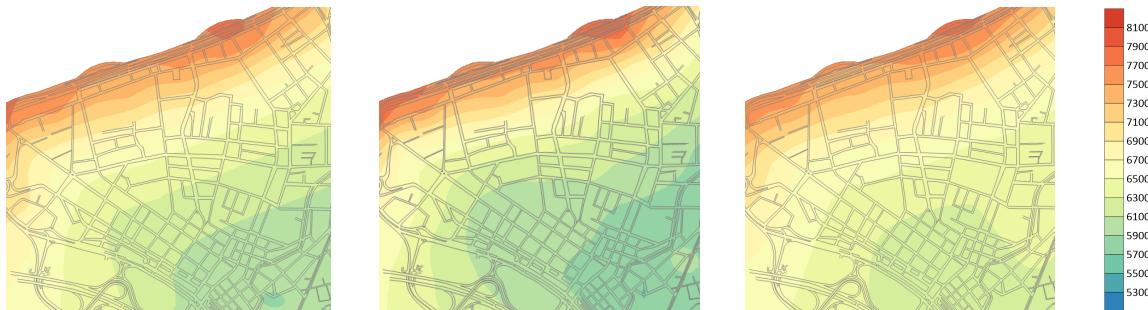


Fonte: o autor

Com relação aos bairros, em Florianópolis, o Centro tem, notadamente, os imóveis mais valorizados. Observa-se, portanto, que a variável localização dos imóveis foi captada de forma muito mais eficiente pela modelagem  $GWR_{SB}$ .

Essa valorização foi muito melhor capturada pela superfície  $GWR_{SB}$ , quando comparada com as superfícies anteriores. Verifica-se que as regiões quentes, que indicam faixas de regiões mais valorizadas, são mostradas de forma mais eficiente na krigagem  $GWR_{SB}$ . Para ilustrar a situação, na Figura 54 é apresentado um recorte da superfície de valores unitários das três modelagens para o bairro Centro.

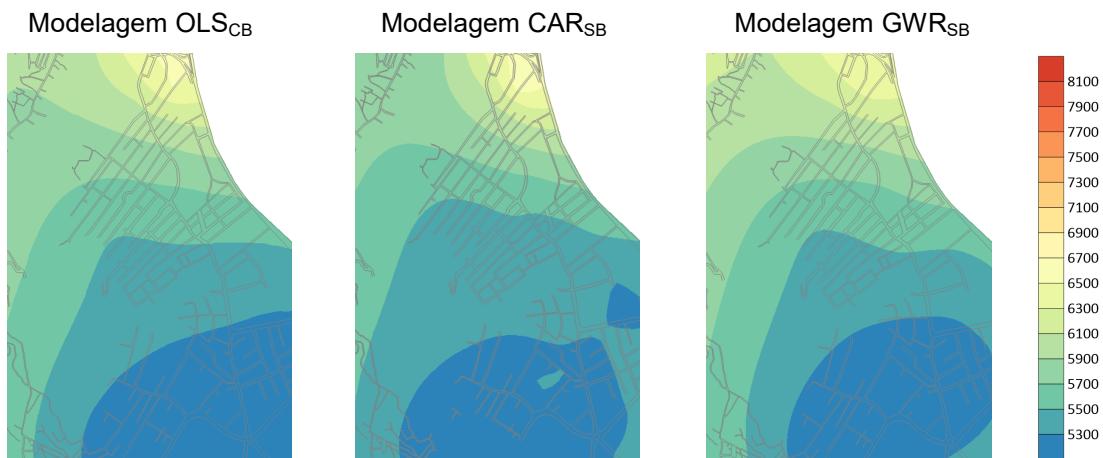
Figura 54: Comparação entre a krigagem das modelagens para o Centro  
 Modelagem OLS<sub>CB</sub>      Modelagem CAR<sub>SB</sub>      Modelagem GWR<sub>SB</sub>



Fonte: o autor

Para o bairro Trindade, observa-se que as três modelagens se apresentaram de maneira semelhante, mostrando imóveis com valor unitário homogeneizado variando entre R\$ 5.100,00 a R\$ 6.300,00 por m<sup>2</sup>. O bairro Trindade, em relação aos outros dois bairros em estudo, possui os imóveis com valor unitário mais baixo.

Figura 55: Comparação entre a krigagem das modelagens para a Trindade



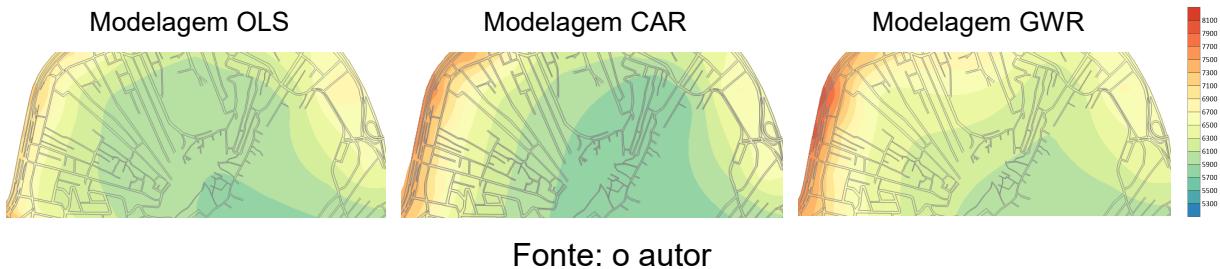
Fonte: o autor

Verifica-se, ainda, que a modelagem do bairro Agronômica apresentou, para a regressão GWR<sub>SB</sub>, faixas que indicam valorização de forma mais eficiente que as outras duas modelagens. Essa característica pode ser verificada na Figura 56.

Em Florianópolis, o bairro Agronômica possui imóveis relativamente novos, de alto padrão e, consequentemente, bem valorizados. É considerado um bairro nobre na região de Florianópolis, bastante cobiçado e com vistas privilegiadas.

Observa-se que a modelagem CAR<sub>SB</sub> também foi capaz de representar eficientemente o valor unitário dos imóveis desse bairro, como visto na Figura 56.

Figura 56: Comparação entre a krigagem das modelagens para a Agronômica



Fonte: o autor

### 6.7.5 Indicadores de Desempenho da Krigagem

Para avaliar o desempenho da krigagem ordinária e verificar qual dos modelos representou de forma mais precisa o mercado imobiliários na região em estudo, foi utilizado, inicialmente, o valor do desvio padrão e do coeficiente de variação da krigagem para as três modelagens. Obteve-se, também, a raiz do valor quadrático médio, do inglês *root mean square* (RMS), conforme a Tabela 42.

Tabela 42: Indicadores de qualidade da krigagem ordinária

Indicador	$OLS_{CB}$	$CAR_{SB}$	$GWR_{SB}$
RMS	6.368,45	6.347,29	6.275,04
Desvio Padrão	719,30	812,35	800,34
Coeficiente de Variação	0,1136	0,1290	0,1266

Fonte: o autor

Constata-se que a krigagem ordinária para os dados preditos pelo modelo de regressão geograficamente ponderada ( $GWR_{SB}$ ) foi a que apresentou o menor valor de RMS, ao passo que a modelagem da regressão clássica ( $OLS_{CB}$ ) foi a que apresentou o menor desvio padrão e o menor coeficiente de variação.

Com relação ao coeficiente de variação, todas as krigagens apresentaram valor inferior a 15%, indicando baixa dispersão (GUJARATI e PORTER, 2011).

Destaca-se que essa análise foi realizada sobre a superfície interpolada com base em 188 dados de mercado, ou seja, sobre o conjunto amostral.

Para complementar a análise da krigagem, realizou-se também o cálculo da raiz quadrada do erro quadrático médio (RMSE) dos valores unitários das 34 amostras utilizadas na etapa de validação dos modelos, utilizando a equação:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (VU_{OBS} - VU_{AJU})^2}{n-1}} \quad (6.3)$$

Onde  $VU$  é o valor unitário e  $n$  é o número de observações da amostra. Nesse caso,  $VU_{OBS}$  são os valores unitários observados nas 34 amostras de validação e  $VU_{AJU}$  são os valores unitários ajustados para as mesmas coordenadas dos  $VU_{OBS}$ . Os  $VU_{AJU}$  foram obtidos para as 34 amostras de validação ao substitui-las na krigagem ajustada sobre os valores unitários homogeneizados e os  $VU_{OBS}$  passaram por homogeneização para que fosse possível o cálculo dos indicadores.

Para cada um dos três modelos gerados por krigagem ordinária calculou-se os valores ajustados ( $VU_{AJU}$ ) a partir dos 34 dados da amostra de validação. Com os valores ajustados para os três modelos de regressão foi possível calcular o RMSE cujos resultados podem ser vistos na Tabela 43.

Tabela 43: Valor do RMSE para krigagem dos valores preditos pelos modelos

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
RMSE (R\$/m <sup>2</sup> )	325,54	318,36	303,77

Fonte: o autor

Constata-se que a krigagem para os dados preditos pelo modelo de regressão geograficamente ponderada (GWR<sub>SB</sub>) foi a que apresentou o menor valor de RMSE para a amostra de validação, indicando ser essa a superfície a apresentar o melhor ajuste dentre as modelagens. Considerando que o menor valor unitário homogeneizado observado é R\$ 5.096,34 e que a RMSE da amostra de validação é inferior a esse valor, pode-se constatar que o modelo ajustado é consistente.

Optou-se, também, pelo cálculo dos parâmetros recomendados pela IAAO (*International Association of Assessing Officers*) para as 34 amostras de validação, a partir do ajuste realizado pela krigagem ordinária, para os três modelos de regressão. Conforme recomendações da IAAO, já indicados no tópico anterior, para os três parâmetros, os limites estão definidos conforme a Tabela 44.

Destaca-se que, nesse estudo, se está avaliando o coeficiente de dispersão COD para imóveis urbanos (apartamentos construídos) em zonas heterogêneas.

Tabela 44: Limites recomendados para os parâmetros de avaliação em massa

Indicador	Recomendação
Mediana das Razões de Avaliação (Med R)	$0,90 < \text{Med R} < 1,10$
Coeficiente de Dispersão (COD)	$\text{COD} \leq 15\%$
Diferencial Relativo ao Preço (PRD)	$0,98 \leq \text{PRD} \leq 1,03$

Fonte: IAAO (2013)

Os resultados dos cálculos dos parâmetros para as 34 amostras de validação ajustadas por krigagem são apresentados na Tabela 45.

Tabela 45: Valores calculados para os parâmetros de avaliação em massa

	<b>OLS<sub>CB</sub></b>	<b>CAR<sub>SB</sub></b>	<b>GWR<sub>SB</sub></b>
Med R	1,004	0,995	0,995
COD (%)	3,29%	4,84%	3,41%
PRD	1,003	1,002	1,001

Fonte: o autor

Com relação à mediana das razões de avaliação (Med R), todos os modelos tiveram seus parâmetros dentro dos limites recomendados pela IAAO.

Quando se analisa o diferencial relativo ao preço (PRD), observa-se que nas três modelagens o valor calculado ficou bastante abaixo do máximo de 1,030 recomendado pela IAAO (2013), o que indica que não houve, nessas modelagens, progressividade ou regressividade na avaliação dos imóveis.

Por fim, com relação à uniformidade horizontal da avaliação obtida por meio do coeficiente de dispersão das razões de avaliação (COD), os valores ficaram abaixo do máximo recomendado pela IAAO para imóveis urbanos, indicando um desempenho bastante elevado para a krigagem das três modelagens realizadas.

Quando comparados esses resultados com aqueles obtidos no Tópico 6.6.2, verifica-se uma melhora ao se utilizar a krigagem. Isso se deve, especialmente, pelo fato de que a krigagem foi realizada sobre os valores homogeneizados, o que acaba por tornar a amostra mais homogênea e, de certa forma, livre de grandes erros.

Contata-se, por fim, que a krigagem para os valores preditos pelo modelo de regressão geograficamente ponderada (GWR) foi a que apresentou os melhores

resultados para a raiz do valor quadrático médio (RMS) e para a raiz do erro quadrático médio (RMSE), mantendo os indicadores Med R, COD e PRD dentro dos limites estabelecidos pela norma IAAO, indicando ser essa a modelagem que, dentre as apresentadas, melhor representaria uma planta de valores para a área em estudo.

## 7 CONSIDERAÇÕES FINAIS

Nos tópicos seguintes serão apresentadas as considerações finais com as conclusões a que se pôde chegar com este estudo, dentro dos limites analisados, bem como algumas sugestões de trabalhos futuros envolvendo os temas aqui abordados. As conclusões sobre alguns pontos da pesquisa são apresentadas a seguir:

### 7.1 CONCLUSÕES

#### **7.1.1 Variáveis do modelo clássico de regressão linear múltipla ( $OLS_{CB}$ )**

Inicialmente, realizou-se a análise exploratória dos dados coletados em meios eletrônicos para, em seguida, se realizar a modelagem clássica. Nesta etapa, analisou-se a matriz de correlações, os diagramas, histogamas e as possíveis transformações que se poderia realizar nas variáveis, objetivando obter modelos consistentes e que não violassem os pressupostos básicos do modelo de regressão.

Obteve-se, então, um modelo de regressão clássica ( $OLS_{SB}$ ) em que as variáveis consideradas mostraram-se significativas e o modelo escolhido não violava os pressupostos básicos, com excessão da dependência espacial nos resíduos.

Como forma de reduzir esses efeitos, optou-se por realizar a modelagem de regressão clássica utilizando variáveis de localização. É importante mencionar que a presença das variáveis indicativas de bairro (variáveis de localização) na modelagem fez com que as estatísticas tivessem seus valores consideravelmente melhorados.

Constatou-se que, ao não se utilizar a variável bairro nas análises, o modelo de regressão clássica apresentava autocorrelação tanto na variável dependente quanto nos termos de erro, indicando, ainda, um possível viés de especificação.

Em vista dos argumentos apresentados, pode-se concluir que a inclusão das variáveis indicativas de bairro na modelagem de regressão clássica ( $OLS_{CB}$ ) fez com que o modelo deixasse de apresentar autocorrelação tanto na variável dependente quanto nos termos de erro, resolvendo, inclusive, um possível viés de especificação.

#### **7.1.2 Matriz de pesos e autocorrelação na regressão espacial ( $CAR_{SB}$ )**

É fato que o modelo clássico de regressão linear é um método consistente e bastante consolidado em engenharia de avaliações, principalmente quando o

objetivo é avaliar um imóvel individualmente. Contudo, ao se realizar avaliações em massa, essa metodologia começa a enfrentar desafios, como o fato de não considerar adequadamente a heterogeneidade e a autocorrelação espacial, que são características comuns em dados de natureza geográfica.

Assim, com auxílio do teste I de Moran, foi possível determinar qual matriz de pesos espaciais conseguia capturar de forma mais efetiva a autocorrelação espacial. Testaram-se diversas matrizes de pesos espaciais, utilizando-se como critérios a distância euclidiana e o número de vizinhos mais próximos. Verificou-se que a matriz mais significativa foi aquela na qual a distância era de 525 metros.

Constatou-se, a partir da matriz de pesos espaciais, que a autocorrelação ocorria nos termos de erro da regressão clássica. Para isso, utilizou-se o teste dos multiplicadores de Lagrange em que apenas o LM do erro mostrou-se estatisticamente significativo. Realizou-se, portanto, a modelagem do erro espacial.

Considerando os aspectos mencionados e os resultados obtidos, conclui-se que o modelo de regressão do erro espacial ( $CAR_{SB}$ ) também foi capaz de controlar a autocorrelação espacial presente nos resíduos da regressão clássica e que o parâmetro de autocorrelação lambda, significativo estatisticamente, permitiu estimar a influência da vizinhança para cada um dos dados da amostra, apresentando-se superior à modelagem OLS<sub>CB</sub> nos indicadores RMSE, ERRO<sub>REL</sub>, COD e PRD.

### **7.1.3 Performance da regressão geograficamente ponderada ( $GWR_{SB}$ )**

Verificou-se, para a regressão geograficamente ponderada ( $GWR_{SB}$ ), as funções de ponderação gaussiana e biquadrática em seus formatos fixo e variável, para diferentes larguras de banda, sendo que o *kernel* gaussiano fixo com 278,42 m de largura de banda foi o que apresentou menor valor no critério de Akaike (AIC).

Escolhida a função de ponderação espacial e a melhor largura de banda, realizou-se o teste DIFF e pode-se verificar que, caso a variável piscina (PS) fosse considerada como variável estacionária, havia uma redução no valor do AIC. Sendo assim, a modelagem  $GWR_{SB}$  foi realizada considerando a variável existência de piscina como global e as demais variáveis do modelo como sendo locais.

Com relação à qualidade, os valores do critério de informação de Akaike (AIC), do critério de informação bayesiano (BIC), da função de verossimilhança (LIK), do coeficiente de determinação ( $R^2$ ), do erro padrão da regressão, da raiz

quadrada do erro quadrático médio (RMSE) e dos erros relativos da regressão geograficamente ponderada mostraram-se superiores às modelagens por regressão clássica e regressão do erro espacial, indicando melhor ajuste e maior aderência.

O desempenho das regressões foi avaliado utilizando uma amostra de validação contendo 34 dados de mercado da área em estudo. Calculou-se, então, qual seria o valor predito para esses dados em cada uma das modelagens e, de posse desses valores, calcularam-se as métricas de qualidade definidas pela IAAO.

Os resultados da mediana das razões de avaliação (Med R), do coeficiente de dispersão horizontal (COD) e do diferencial relativo ao preço (PRD) mostraram-se dentro dos limites estabelecidos pela IAAO para as três modelagens. Deve-se destacar a modelagem GWR<sub>SB</sub> apresentou o menor valor para o COD, indicando menor dispersão horizontal das razões de avaliação, e valor de PRD próximo à unidade, quando comparado à modelagem clássica e do erro espacial.

Calculou-se, também, para os valores preditos para amostra de validação das três modelagens, a raiz quadrada do erro quadrático médio (RMSE) e os erros relativos (absolutos) de cada ponto de validação. Os resultados indicaram menores valores para o RMSE e para os erros relativos na regressão GWR<sub>SB</sub>, indicando ser essa a melhor modelagem para os dados de mercado da área em estudo.

Dado o exposto, pode-se concluir que a regressão GWR<sub>SB</sub> foi capaz de controlar de forma mais eficiente os efeitos espaciais da modelagem OLS<sub>SB</sub> que a simples inclusão das variáveis de localização (bairros) na regressão clássica ou que a modelagem por regressão espacial global. Além disso, a modelagem GWR<sub>SB</sub> mostrou-se superior em todos os indicadores de qualidade e desempenho, quando comparado aos valores obtidos nas modelagens clássica e espacial global.

#### **7.1.4 Interpolação dos valores unitários por krigagem ordinária**

Definiu-se o imóvel paradigma que serviu de referência para a homogeneização de todos os dados da amostra. Construíram-se, então, mapas variográficos para as três modelagens e pôde-se constatar que não havia mudanças de gradientes nas diversas direções, permitindo concluir que o fenômeno em estudo era isotrópico.

Nesse sentido, um único semivariograma teórico pode ser representativo da estrutura de variação espacial da variável avaliada. Construiu-se, então, o semivariograma omnidirecional experimental para cada uma das modelagens e constatou-se

que o semivariograma teórico exponencial era o que melhor se ajustava aos dados obtidos pelas regressões. Esse modelo teórico mostrou que a maior parte da variabilidade dos resíduos é atribuída a autocorrelação espacial.

O processo de validação cruzada permitiu confirmar que o semivariograma teórico exponencial era o adequado para os dados, pois apresentou maior valor para o coeficiente de determinação e menor valor para a raiz do erro quadrático médio.

Com isso foi possível realizar a krigagem ordinária para gerar a superfície de valores unitários homogeneizados preditos pelas três modelagens e pela superfície de erro da krigagem. Para regressão GWR<sub>SB</sub>, a homogeneização dos valores unitários de cada ponto amostral se deu sobre a regressão do próprio ponto.

Verificou-se que a superfície de valores unitários gerada para a modelagem GWR<sub>SB</sub> foi capaz de representar de maneira mais eficiente, com resultados mais próximos do esperado, o valor unitário dos imóveis nos bairros em estudo.

Pode-se observar, também, que a krigagem do modelo OLS<sub>CB</sub> apresentou problemas de estimação para o bairro Agronômica, ao passo que o modelo CAR<sub>SB</sub> apresentou problemas de estimação no bairro Centro. Já a modelagem GWR<sub>SB</sub> conseguiu representar adequadamente o valor unitário dos imóveis nesses bairros.

O desempenho das três superfícies de valores unitários homogeneizados geradas por krigagem foram avaliadas na predição de 34 dados da amostra de validação. Os resultados mostraram que os valores obtidos para as três modelagens se mantiveram dentro dos limites estabelecidos pela IAAO, contudo, a modelagem GWR<sub>SB</sub> apresentou o valor do indicador de equidade vertical (PRD) mais próximo da unidade e o menor valor para a raiz quadrada do erro quadrático médio (RMSE).

Por todos esses aspectos levantados, pode-se concluir que a superfície de valores unitários, gerada pela modelagem GWR<sub>SB</sub>, mostrou-se mais aderente, com valores preditos mais próximos dos valores praticados no mercado, representando de forma mais eficiente a realidade do mercado imobiliário da área em estudo.

### **7.1.5 Avaliação em massa de apartamentos da região estudada**

O estudo efetuado permitiu tratar de forma eficiente os efeitos espaciais de três formas distintas (inclusão das variáveis de localização na regressão clássica, uso da regressão do erro espacial e uso da regressão geograficamente ponderada).

Contudo, a modelagem GWR<sub>SB</sub> se mostrou superior às modelagens globais (OLS<sub>CB</sub> e CAR<sub>SB</sub>) em todos os indicadores. Destaca-se que, nas superfícies de erros

relativos geradas, a modelagem GWR<sub>SB</sub> apresentou-se mais uniforme, com erros de menor magnitude. Para avaliações em massa isso é desejável, o que demonstra que a parcela de erro foi aproximadamente constante em toda a área em estudo.

Em vista dos argumentos apresentados, conclui-se que a modelagem GWR<sub>SB</sub> mostrou-se a mais adequada para a avaliação em massa de imóveis urbanos.

Como foram utilizados valores homogeneizados para a geração da superfície de valores da modelagem GWR, esse produto cartográfico é utilizado para indicar o valor genérico do metro quadrado dos apartamentos da área em estudo.

Frente a isso, o método proposto pode ser de grande utilidade para os órgãos municipais responsáveis pelo cálculo e pela atualização de produtos cartográficos que utilizam avaliações em massa em sua composição, como a PVG.

## 7.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Tendo como base os resultados e as conclusões alcançadas neste trabalho, são apresentadas a seguir algumas recomendações para futuras pesquisas:

1. Investigar o processo de avaliação dos dados, utilizando funções de ponderação adaptativas, baseadas em número de vizinhos mais próximos, para a regressão geograficamente ponderada. Neste estudo foi adotada a função de ponderação gaussiana fixa, tendo como base a distância do ponto de regressão aos vizinhos, contudo, pesquisas que considerem o número de vizinhos podem ser interessantes;
2. Investigar como se daria o uso da regressão geograficamente ponderada (GWR) para a avaliação em massa de apartamentos de determinada área, abrangendo um número maior de bairros ou, inclusive, a área referente a um município inteiro;
3. Investigar como se daria o uso desta metodologia (GWR) para a avaliação de outros tipos de imóveis urbanos, como casas, salas comerciais ou terrenos.

## REFERÊNCIAS

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-1: Avaliação de Bens.** Parte 1: Procedimentos Gerais. Rio de Janeiro, 2019. 12 p.

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2: Avaliação de Bens.** Parte 2: Imóveis Urbanos. Rio de Janeiro, 2011. 53 p.

AFONSO, J. R.; ARAUJO, E. A. e NÓBREGA, M. A. R. **IPTU no Brasil:** um diagnóstico abrangente. v. 4. São Paulo: FGV Projetos, 2014.

AKAIKE, H. **A new look at the statistical model identification.** IESS transactions on automatic control. Boston. v. 19, n. 6, p. 716-723, 1974.

ALMEIDA, Eduardo. **Econometria espacial aplicada.** São Paulo: Alínea, 2012.

ANSELIN, L. **Exploring spatial data with GeoDaTM:** A Workbook. Department of Geography, University of Illinois, 2005.

ANSELIN, L. **Spatial econometrics:** discussion paper. Bruton Center, School of Social Sciences, University of Texas at Dallas, 1999.

ANSELIN, L. **Spatial econometrics:** methods and models. Kluwer Academic Publishers, 1988.

ANSELIN, L. **The moran scatterplot as an ESDA tool to asses local instability in spatial association.** In: Spatial Analytical Perspectives on GIS. London: Taylor and Francis. p. 111-125, 1996.

ARBIA, Giuseppe. **A primer for spatial econometrics:** with applications in R. AIAA. 2014, 248 p.

AXHAUSEN, K. W.; LOCHL, M. (2010). **Modeling hedonic residential rents for land use and transport simulation while considering spatial effects.** Journal of Transport. v. 3, n. 2, p. 39-63, 2010.

BALEIRO, Aliomar e DERZI, Misabel. **Direito tributário brasileiro:** CTN comentado. 13. ed. Rio Janeiro: Forense, 2016.

BIDANSET, P. E.; LOMBARD, J. R. (2014) **Evaluating spatial model accuracy in mass real estate appraisal:** a comparison of geographically weighted regression and the spatial lag model. Cityscape. v. 16, n. 3, 2014.

BOX, G. E. P.; COX, D. R. (1964). **An analysis of transformations.** Journal of the royal statistical society. Series B, v. 26, p. 211-252, 1964.

BOZDOGAN, H. (1987). **Model selection and Akaike's Information Criterion (AIC):** the general theory and its analytical extensions. Psychometrika. v. 52, n. 3, p. 345–370, 1987.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil.** Brasília: DF, 1988.

BRASIL. **Código Tributário Nacional - CTN.** Lei 5.172 de 25 de outubro de 1966. Brasília: DF, 1966.

BRUNSDON C.; FOTHERINGHAM A. S.; CHARLTON M. (1999). **Some notes on parametric significance tests for geographically weighted regression.** Journal of Regional Science, v. 39, p. 497–524, 1999.

BRUNSDON, C.; FOTHERINGHAM, A. S.; e CHARLTON, M. E. (1996). **Geographically weighted regression:** a method for exploring spatial nonstationarity. Geographical Analysis, v. 28, p. 281-298, 1996.

BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference: a practical information-theoretic approach.** New York: Springer, 2002, 488 p.

CÂMARA, G.; MONTEIRO, A. M.; CARVALHO, M. S.; DRUCK, S. (2002). **Análise espacial de dados geográficos.** DPI - INPE. 2002.

CAMARGO, E. C. G.; FUCKS, S. D.; CAMARA, G. **Análise espacial de superfícies.** In: análise espacial de dados geográficos (Capítulo 3). INPE. Instituto Nacional de Pesquisas Espacial. 2004.

CARVALHO, L. E. X.; SILVA, H. N.; LOUREIRO, C. F. G.; MENESSES, H. B. Regressão linear geograficamente ponderada em ambiente SIG. **Revista Transportes.** v. XIV, n. 2, p. 18-26, dez. 2006.

CARVALHO JR, P. H. B. **O sistema avaliatório municipal de imóveis e a tributação do IPTU no Rio de Janeiro.** 2011. 101 f. Dissertação (Mestrado em Ciências Econômicas) – Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2011.

CELLMER, Radosław. (2012). **The use of the geographically weighted regression for the real estate market analysis.** Folia Oeconomica Stetinensis. 2012.

CESARE, C. M. de. **An empirical analysis of equity in property taxation:** a case study from Brazil. Salford. 1998. Tese de Doutorado, University of Salford – UK.

CESARE, C. M. de; **Avaliação em massa de imóveis para fins fiscais:** discussão, análise e identificação de soluções para problemas e casos práticos. In: programa nacional de capacitação das cidades e Lincoln Institute of Land Policy. Org. CUNHA, E. P. C.; ERBA, D. A. Brasília: Ministério das Cidades, 2012. 116 p.

CESARE, C. M. de; AVERBECK, C. E. **Da avaliação de imóveis.** In: manual de apoio CTM: diretrizes para a criação, instituição e atualização do cadastro territorial multifinalitário nos municípios brasileiros. Org. CUNHA, E. P. C.; ERBA, D. A. Brasília: Ministério das Cidades, 2010. 170 p.

CHARLTON, M.; FOTHERINGHAM, S.; BRUNSDON, C. (2009). **Geographically weighted regression**. White Paper. National Centre for Geocomputation. National University of Ireland Maynooth. 2009.

CLEVELAND, W. S. **Robust locally weighted regression and smoothing scatterplots**. Journal of the american statistical association, v. 74, p. 829-836, 1979.

DANTAS, R. A. **Modelos espaciais aplicados ao mercado habitacional**: um estudo de caso para a cidade de Recife. 2003. 133 f. Tese (Doutorado em Economia) – Universidade Federal de Pernambuco (UFPE), Recife, 2003.

DANTAS, R. A.; ROCHA, F. J. S.; MAGALHÃES, A. M.; LIMA, R. C. Uma nova metodologia para avaliação de imóveis utilizando regressão espacial. In: Congresso Brasileiro de Engenharia de Avaliações e Perícias, XI, 2001, Guarapari. **Anais...** Espírito Santo, 2001.

DANTAS, Rubens Alves. **Engenharia de avaliações**: uma introdução à metodologia científica. 3. ed. São Paulo: Pini, 2012. 255 p.

DIMOPOULOS, T.; MOULAS, A. (2016). **A proposal of a mass appraisal system in Greece with CAMA system**: evaluating GWR and MRA techniques in Thessaloniki municipality. Gruyter Open. 2016.

DUALDE, R. **IPTU sob medida**: análise de dados fiscais e socioeconômicos para a tributação de imóveis residenciais. 2004. 183 f. Dissertação (Mestrado em Estruturas Ambientais Urbanas) – Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo (FAU/USP), São Paulo, 2004.

DUBIN, R. **Robustness of spatial autocorrelation specifications**: some Monte Carlo evidence. Journal of regional science. v. 43, p. 221-248, 2003.

DZIAUDDIN, M. F., IDRIS, Z. (2017). **Use of geographically weighted regression (GWR)**: method to estimate the effects of location attributes on the residential property values. Indonesian Journal of Geography. v. 49, n. 1, Indonesian, p. 97-110, 2017.

FARIA FILHO, R. F. **Modelos estatísticos para geração de plantas de valores genéricos em áreas urbanas**. 2015. 129 f. Tese (Doutorado em Geografia) – Universidade Federal de Uberlândia (UFU), Uberlândia. 2015.

FLORES, R. M.; GARCIA, M. J. S. (2018). **El componente espacial en la valoración de inmuebles residenciales**: uso de la regresión geográficamente ponderada. Geografía y Sistemas de Información Geográfica (GEOSIG). Luján, a. 10, n. 10, p. 167-188. 2018.

FONTOURA Jr., C. F. M.; UBERTI, M. S. Avaliação em massa de imóveis urbanos utilizando regressão múltipla e estimador kernel. In: Congresso Brasileiro de Cadastro Multifinalitário, XIII, 2018, Florianópolis. **Anais...** Santa Catarina, 2018.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. (1998). **Geographically weighted regression:** a natural evolution of the expansion method for spatial data analysis. Environment and Planning, v. 30, p. 1905-1927, 1998.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. (2000). **Quantitative geography:** perspectives on spatial data analysis. London: Sage, 2000.

FOTHERINGHAM, A.; BRUNSDON, C.; CHARLTON, M. (2002). **Geographically weighted regression:** the analysis of spatially varying relationship. Chichester. 2002.

FOTHERINGHAM, A.S.; CHARLTON, M.; BRUSDON, C. (1997). **Recent developments in spatial analysis:** chapter measuring spatial variations in relationships with geographically weighted regression. Springer, p. 60-82. New York, EUA, 1997.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa.** Porto Alegre: UFRGS, 2009.

GIL, A. C. **Como elaborar projetos de pesquisas.** 7. ed. São Paulo: Atlas, 2017.

GONZÁLEZ, M. A. S. **A engenharia de avaliações na visão inferencial.** São Leopoldo: Unisinos, 1997.

GONZÁLEZ, M. A. S. **Aplicação de técnicas de descobrimento de conhecimento em base de dados e de inteligência artificial em avaliações de imóveis.** 2002. 300 f. Tese (Doutorado em Engenharia Civil) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

GOVERNO de Santa Catarina. **Município de Florianópolis,** Página Inicial. Disponível em: <<http://sol.sc.gov.br>>. Acesso em: 05 nov. 2019.

GUIMARÃES, E. C. **Geoestatística básica e aplicada.** Uberlândia, MG. 2004.

GUJARATI, D. N; PORTER, D. C. **Econometria básica.** 5. ed. Porto Alegre: AMGH Bookman, 2011.

HORNBURG, Ricardo André. **Avaliação em massa com uso combinado da regressão espacial e da geoestatística:** estudo de caso em Navegantes/SC. 2009. 79 f. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2009.

HORNBURG, Ricardo André. **Avaliação em massa de imóveis usando estatística bayesiana.** 2015. 138 f. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2015.

IAAO (International Association of Assessing Officers). **Standards on ratio studies.** Mirrouri: IAAO. 2013.

JAHANSIRI, E.; BUYONG, T.; SHARIFF, A. R. M. (2011). **A review of property mass valuation models**. Pertanika J. Sci. & Technol. Malaysia, p. 23-30, 2011

KREMPI, A. P. **Explorando recursos de estatística espacial para análise da acessibilidade na cidade de Bauru**. 2004. 94 f. Dissertação (Mestrado em Engenharia) – Escola de Engenharia de São Carlos. Universidade de São Paulo, São Carlos, 2004.

LAI, P. **Analysis of the mass appraisal model using artificial neural network in Kaohsiung city**. Journal of Modern Accounting and Auditing. October 2011. v. 7, n. 10, New York, p. 1080-1089, 2011.

LEG, R.; BOWE, T. (2009). **Applying geographically weighted regression to a real estate problem**: an example from Marquette, Michigan. ArcUser Spring. 2009.

LEUNG Y.; MEI C. L.; ZHANG W. X. (2000). **Statistical tests for spatial nonstationarity based on the geographically weighted regression model**. Environment and Planning, v. 32, p. 9–32, 2000.

LIMA, O. P.; CORDINI, J.; LOCH, C. O cadastro técnico multifinalitário e o poder público municipal: base para o desenvolvimento sustentável. In: Congresso Brasileiro de Cadastro Técnico Multifinalitário, IV, 2000, Florianópolis. Anais... Santa Catarina, 2000.

LIPORONI, A. S. **Avaliação em massa com ênfase em planta de valores**. In: ALONSO, Nelson Roberto Pereira (Coord). Engenharia de Avaliações. 2. ed. São Paulo: PINI, 2013.

LOCKWOOD, T.; ROSSINI, P. (2012). **Efficacy of the geographically weighted model on the mass appraisal process**. 18 th. Annual Pacific-RIM Real State Society Conference. Adelaide, Austrália. 2012.

LOUREIRO, C.F.G.; SILVA, H. N.; CARVALHO, L. E. X. Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. In: Congresso de Pesquisa e Ensino em Transportes, XX, 2006, Brasília. Anais... Distrito Federal, 2006.

MALAMAN, C. S. **Aplicação de lógica fuzzy na elaboração de planta de valores genéricos**. 2014. 86 f. Dissertação (Mestrado em Ciências Cartográficas) – Universidade Estadual Paulista (UNESP), São Paulo, 2014.

MARQUES, A. P. S.; JUNIOR, J. M.; AMORIM, A.; FLORES, E. F. Aplicação do interpolador krigagem ordinária para elaboração de planta de valores genéricos. **Revista Brasileira de Cartografia**. v. 64, n. 2, p. 175-186, 22 maio 2012.

McCLUSKEY, W. J.; DEDDIS, W. G.; MANNIS, A.; McBURNEY, D.; BORST, R. A. Interactive application of computer assisted mass appraisal and geographic information systems. **Journal of Property Valuation and Investment**. v. 15, n. 5, p. 448-465, 1997.

MEDINA, F. A. S. **Regressão logística geograficamente ponderada aplicada a modelos de credit scoring.** 2016. 93 f. Dissertação (Mestrado em Administração) – Universidade de Brasília (UNB), Distrito Federal, 2016.

MEI, Chang-Lin. **Geographically weighted regression technique for spatial data analysis.** School of Science. ResearchGate. 2004, 14 p.

MELLO, Celso Antônio Bandeira de. **Curso de direito administrativo.** 33. ed. São Paulo: Malheiros, 2015, p. 382.

MENDONÇA, Marcelo Corrêa *et al.* **Fundamentos de avaliações patrimoniais e perícias de engenharia:** curso básico do IMAPE. São Paulo: Pini, 1998.

MÖLLER, L. F. C. **Planta de valores genéricos:** avaliação coletiva de imóveis para fins tributários. Porto Alegre: Sagra-DC Luzzatto, 1995.

MOURA, E. M.; CARNEIRO, A. F. T. Planta de valores para municípios de pequeno porte: o caso de Salgadinho (PE). In: Simpósio em Ciência Geodésicas e Tecnologias da Geoinformação, I, 2004, Recife. **Anais...** Pernambuco, 2004.

NAKAYA, Tomoki. **Manual de uso GWR4.** Windows application for geographically weighted regression modelling. 2016.

RADEGAZ, Nasser Júnior. **Avaliação de bens:** princípios básicos e aplicações. 2. ed. São Paulo: LEUD, 2013. 224 p.

RAMOS, L. S.; OLIVEIRA, R.; LOCH, C.; SILVA, E. Banco de dados do mercado imobiliário integrado ao cadastro técnico multifinalitário: formação de técnicos em plantas de valores genéricos (PVG). In: Congresso Brasileiro de Cadastro Técnico Multifinalitário, V, 2002, Florianópolis. **Anais...** Santa Catarina, 2002.

RUDIO, F. V. **Introdução ao projeto de pesquisa científica.** 9. ed. Petrópolis: Vozes, 1985.

SA, J. D. **Um modelo para o IPTU dos imóveis residenciais que visa incentivar a sustentabilidade urbana.** 2014. 268 f. Tese (Doutorado em Engenharia Industrial) - Universidade Federal da Bahia, Salvador, 2014.

SCHIAVO, E. H. C.; AZEVEDO, M. P. Estudo comparativo entre redes neurais artificiais e análise de regressão múltipla na avaliação de bens para pequenas amostras. In: Congresso Brasileiro de Engenharia de Avaliações e Perícias, 2003, Belo Horizonte. **Anais...** Minas Gerais, 2003.

SCHWARZ, G. **Estimating the dimensional of a model.** Annals of statistics. Hayward, v. 6, n. 2, p. 461-464, 1978.

SECRETARIA DE TURISMO. **História de Florianópolis.** Página Inicial. Disponível em: <<http://www.pmf.sc.gov.br/entidades/turismo>>. Acesso em: 05 nov. 2019.

SERBANOIU, Ion Ventorim (Org). **Planta de valores imobiliários:** relatório sobre a modernização da metodologia de avaliação de imóveis em Belo Horizonte. Belo Horizonte: Prefeitura, 2019. Disponível em: <<https://prefeitura.pbh.gov.br/fazenda/simposio-modelagem-de-avaliacao-de-imoveis>>. Acesso em: 05 nov. 2019.

SERRANO, R. M.; VALCARCE, E. V. **Técnicas econométricas para el tratamiento de datos espaciales:** la econométrica espacial. Edicions Universita de Barcelona, Barcelona, 2000.

SILVA, A. C. da C. **Análise do problema da unidade de área modificável pela regressão geograficamente ponderada.** 2019. 119 f. Dissertação (Mestrado em Estatística) – Universidade de Brasília (UNB), Distrito Federal, 2019.

SILVA, A. R. **Metodologia para avaliação e distribuição de recursos para o transporte escolar rural.** 2009. 161 f. Tese (Doutorado em Engenharia Civil e Ambiental) – Universidade de Brasília (UNB), Distrito Federal, 2009.

SUNAK, Y.; MADLENER, R. (2012). **The impact of wind farms on property values:** a geographically weighted hedonic pricing model. FCN Working, v. 3, 2012.

THEODORO, L. T. C.; UBERTI, M. S.; ANTUNES, M. A. H.; DEBIASI, P. Avaliação em massa de imóveis rurais através da regressão clássica e da geoestatística. **Revista Brasileira de Cartografia.** v. 71, n. 2, p. 459-485, 24 jun. 2019.

THOFEHRN, Ragnar. **Avaliação de terrenos urbanos por fórmulas matemáticas.** São Paulo: Pini, 2008.

TIRYAKI, G. F.; ANDRADE, C. S. M. **Econometria na prática.** 1 ed. São Paulo: Alta Books, 2017.

TOBLER, W. R. **A computer movie simulating urban growth in the Detroit region.** Economic Geography. v. 46, p. 234-240, 1970.

TRISTAO, J. A. M. **A Administração tributária dos municípios brasileiros:** Uma avaliação do desempenho da arrecadação. 2003. 172 f. Tese (Doutorado em Organização, Recursos Humanos e Planejamento.) EAESP/FGV, São Paulo, 2003.

TRIVELLONI, C. A. P. **Método para determinação do valor da localização com uso de técnicas inferenciais e geoestatísticas na avaliação em massa de imóveis.** 2005. 172 f. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2005.

TRIVELLONI, C. A. P., HOCHHEIM, N. O valor de localização dos imóveis: determinação por métodos de análise espacial. In: Congresso Brasileiro de Engenharia de Avaliações e Perícias, XIII, 2006, Fortaleza. **Anais...** Ceará, 2006.

TYSZLER, M. **Econometria espacial:** discutindo medidas para a matriz de ponderação espacial. 2006. 141 f. Dissertação (Doutorado em Administração Pública) – Fundação Getúlio Vargas (FGV), São Paulo, 2006.

UBERTI, M. S. **Avaliação em massa de imóveis rurais através de modelagem clássica, espacial e geoestatística.** 2016. 151 f. Tese (Doutorado em Ciência, Tecnologia e Inovação) – Universidade Federal Rural do Rio de Janeiro (UFRRJ), Rio de Janeiro, 2016.

VASCONCELLOS, M. V. Reflexões sobre cadastros da América Latina. *In:* Seminário Internacional de Tributação Imobiliária, 2001. **Anais...** Porto Alegre, 2001.

VERGARA, S. C. **Métodos de pesquisa em administração.** 6 ed. São Paulo: Editora Atlas, 2015.

VIEIRA, R. S. **Crescimento econômico no estado de São Paulo: uma análise espacial [online].** São Paulo: Cultura Acadêmica, 2009.

WYATT, P. **The development of a property information system for valuation using a Geographical Information System (GIS).** Journal of Property Research. v. 13, p. 317-336, 1996.

## APÊNDICE 01: AMOSTRA DE DADOS DE MERCADO (TRABALHO)

NOME	VT	VU	AP	DPXV	DSBM	DSIG	DCTC	DABM	ND	NB	NS	NG	MO	PS	CH	PC	BRO	COORD_E	COORD_N
AP_01	920045	9485	97	1642	876	6763	7022	367	3	3	1	1	MO	S	S	A	Centro	742452	6946217
AP_02	699997	9589	73	2033	1266	6916	7175	807	2	3	2	2	N	S	S	M	Centro	742745	6945881
AP_03	750000	6250	120	1122	1324	7211	7470	803	2	4	2	2	N	N	N	M	Centro	742245	6945636
AP_04	1149972	9426	122	1612	1064	6952	7211	136	3	3	1	1	SM	N	N	A	Centro	741795	6946170
AP_05	700040	6364	110	1532	936	6824	7083	116	3	2	1	1	MO	N	N	M	Centro	741679	6946144
AP_06	630000	7500	84	2376	1668	7543	7802	143	2	2	0	1	N	N	N	B	Centro	741383	6946018
AP_07	839990	8842	95	2480	2615	8493	8752	319	3	2	1	1	MO	S	S	M	Centro	741004	6945712
AP_08	530010	9815	54	1371	1971	7846	8105	892	3	1	0	1	MO	N	N	A	Centro	741606	6945303
AP_09	1075020	7790	138	1517	1291	7166	7425	585	3	4	1	2	MO	N	N	M	Centro	742133	6945825
AP_10	529980	8833	60	1372	1164	6982	7241	660	2	2	1	1	MO	N	N	M	Centro	742338	6945830
AP_11	1255104	13074	96	2055	1269	7156	7415	282	4	3	1	2	N	S	N	A	Centro	742205	6946185
AP_12	849960	7083	120	1119	1349	7237	7496	475	3	3	2	2	MO	S	N	M	Centro	741652	6945754
AP_13	990000	8250	120	1617	1427	7302	7561	338	3	3	1	1	MO	S	S	A	Centro	741479	6945844
AP_14	2399957	14371	167	1620	834	6721	6980	528	4	5	3	3	MO	S	S	A	Centro	742456	6946030
AP_15	458012	6836	67	1828	1042	6929	7188	358	2	1	1	1	SM	N	N	M	Centro	742278	6946132
AP_16	889985	8165	109	1373	1497	7450	7709	674	3	3	1	2	N	S	S	M	Centro	742086	6945709
AP_17	1150050	10455	110	2401	621	6508	6767	91	3	4	1	1	SM	N	N	A	Centro	741932	6946277
AP_18	1849965	11635	159	2989	2282	8160	8419	330	3	4	1	2	SM	S	S	A	Centro	740879	6945635
AP_19	1179957	9291	127	1647	1099	6987	7246	140	3	4	1	2	N	N	N	M	Centro	741826	6946180
AP_20	639979	8101	79	2508	2643	8521	8780	299	3	2	1	1	N	N	S	A	Centro	741041	6945746
AP_21	235008	4352	54	364	2494	7642	4609	1425	2	1	0	0	MO	N	N	B	Centro	742066	6944890
AP_22	530029	8689	61	1439	1275	7262	7521	552	2	2	1	1	SM	N	N	M	Centro	742277	6945921
AP_23	1248000	13000	96	3074	2366	8244	8503	98	3	3	1	1	MO	S	S	A	Centro	740979	6945937
AP_24	649029	9687	67	470	2076	8456	4740	1244	2	2	1	1	SM	S	S	M	Centro	742236	6945154
AP_25	360006	4138	87	844	2306	7866	4560	1207	2	2	0	1	SM	N	N	B	Centro	742432	6945276
AP_26	2447025	8295	295	3094	2387	8265	8524	411	4	4	2	3	N	S	S	A	Centro	740966	6945573
AP_27	1274014	12614	101	1132	1423	7311	7570	536	3	2	1	2	N	S	S	A	Centro	741739	6945717
AP_28	549976	9016	61	2382	2517	8395	8654	420	2	1	0	1	N	N	N	M	Centro	741043	6945619
AP_29	449972	6716	67	1691	905	6792	7051	490	3	2	1	1	MO	N	N	M	Centro	742373	6946029
AP_30	1649928	10248	161	1597	1408	7283	7542	310	3	4	1	2	SM	N	S	A	Centro	741490	6945876
AP_31	536983	8803	61	2547	2682	8560	8819	266	2	1	0	1	MO	N	N	M	Centro	741048	6945784
AP_32	784000	7840	100	1337	2178	7746	5236	625	3	4	2	1	SM	N	S	M	Centro	742501	6945944
AP_33	915048	6444	142	2128	2997	8870	9129	321	3	3	1	2	N	N	S	B	Centro	740883	6945477
AP_34	1039969	9541	109	1471	1595	7547	7806	625	3	2	1	2	MO	N	S	A	Centro	741925	6945695
AP_35	1200000	8000	150	1993	1803	7678	7937	353	3	4	1	2	SM	N	N	M	Centro	741252	6945758
AP_36	1149952	8984	128	2618	1910	7788	8047	74	3	2	1	2	MO	N	N	A	Centro	740698	6945861
AP_37	1300000	6500	200	930	1594	7482	7741	700	3	4	3	2	N	N	N	A	Centro	741731	6945543
AP_38	1889927	8957	211	2277	459	6346	6605	91	4	4	2	2	N	S	N	A	Centro	742073	6946336
AP_39	625986	7542	83	2580	2716	8594	8853	235	3	2	1	1	MO	S	S	M	Centro	741053	6945818
AP_40	1049040	11160	94	2967	2259	8137	8396	220	3	2	1	2	N	N	S	A	Centro	740831	6945761
AP_41	1949968	14338	136	3690	1625	7503	7762	153	4	4	3	2	MO	S	S	A	Centro	741049	6945903
AP_42	349988	5932	59	2492	2652	9139	9398	633	2	1	0	1	N	N	S	B	Centro	741193	6945430
AP_43	315000	5625	56	1415	2195	7336	4221	1010	1	2	1	0	N	N	S	A	Centro	742666	6945595
AP_44	680002	7907	86	2062	1872	7747	8006	395	3	2	1	1	SM	N	S	M	Centro	741175	6945689
AP_45	1049991	9813	107	2426	2561	8439	8698	357	3	2	1	3	SM	S	N	A	Centro	740987	6945666
AP_46	830032	7411	112	1135	1917	7805	8064	548	3	2	1	1	SM	N	N	A	Centro	741762	6945713
AP_47	580545	5985	97	3877	1811	7689	7948	345	2	2	1	1	SM	N	N	B	Centro	741143	6945732
AP_48	830025	9765	85	1571	1381	7256	7515	302	3	2	1	2	N	S	N	M	Centro	741549	6945903
AP_49	579020	8515	68	1443	2223	7364	4249	997	2	2	1	2	N	S	S	A	Centro	742628	6945589
AP_50	729968	8488	86	2196	2705	9582	9841	544	3	2	1	1	MO	S	S	M	Centro	741042	6945236
AP_51	855998	11726	73	1799	1013	6900	7159	637	3	3	2	2	N	S	S	A	Centro	742435	6945895

AP_52	599040	6240	96	698	1855	7452	4688	1085	3	2	1	1	N	S	S	B	Centro	742301	6945353
AP_53	597984	6229	96	2835	2127	8005	8264	151	2	2	1	1	MO	S	S	B	Centro	740782	6945819
AP_54	558000	6200	90	2193	2702	9579	9838	514	2	1	1	1	N	S	S	M	Centro	741024	6945266
AP_55	1060008	12184	87	2086	1300	7187	7446	265	3	2	1	2	SM	N	S	A	Centro	742130	6946171
AP_56	2799940	12727	220	1280	2062	7950	8209	477	4	5	3	4	N	S	S	A	Centro	741975	6945876
AP_57	749994	5102	147	2801	2093	7971	8230	143	2	3	1	1	MO	N	N	M	Centro	740741	6945806
AP_58	1102620	11730	94	1853	1067	6954	7213	668	3	3	2	2	N	S	S	A	Centro	742390	6945843
AP_59	730020	10580	69	2428	2427	8305	8564	592	2	2	1	1	MO	N	S	A	Centro	741469	6945575
AP_60	349980	5833	60	1768	982	6869	7128	400	3	1	0	1	N	N	N	B	Centro	742326	6946106
AP_61	953232	11348	84	556	2231	8119	8378	1229	3	2	1	1	N	S	S	A	Centro	741906	6945043
AP_62	849990	4885	174	549	2308	8196	4819	967	3	5	2	2	N	N	N	B	Centro	741827	6945293
AP_63	1290060	9556	135	1309	2091	7979	8238	504	3	4	2	2	N	S	S	M	Centro	741973	6945846
AP_64	849986	7522	113	1391	1515	7467	7726	633	3	3	1	2	SM	N	S	M	Centro	742002	6945718
AP_65	679995	9855	69	2137	3006	8879	9138	113	2	2	1	1	SM	N	N	M	Centro	742438	6946556
AP_66	750040	8824	85	1648	1142	7203	7462	465	1	2	1	1	MO	S	S	A	Centro	742245	6946002
AP_67	967694	10634	91	2245	620	6307	6566	237	3	2	1	2	MO	S	S	A	Centro	742567	6946529
AP_68	1200040	7895	152	1987	855	6394	6653	328	3	2	1	2	SM	S	S	B	Centro	742570	6946364
AP_69	649000	5900	110	2187	1733	6873	7132	853	2	3	1	1	MO	N	S	B	Centro	742538	6945705
AP_70	560031	7089	79	618	2748	5429	4352	1362	2	1	0	1	MO	N	S	M	Centro	742278	6945043
AP_71	1300043	11927	109	926	1932	7820	8079	906	3	3	1	2	MO	N	S	A	Centro	742018	6945429
AP_72	289018	4738	61	1685	3387	9260	9519	801	2	1	0	0	N	N	N	B	Centro	741327	6945254
AP_73	900030	4737	190	861	1815	7703	7962	768	3	4	1	2	SM	N	N	B	Centro	741617	6945436
AP_74	649977	6989	93	2114	2000	7878	8137	498	2	2	1	2	SM	N	N	B	Centro	741183	6945584
AP_75	2645030	15559	170	1361	1741	7629	7888	740	4	4	3	4	MO	S	S	A	Centro	741909	6945563
AP_76	1390066	9521	146	2365	2364	8242	8501	716	4	3	1	2	MO	N	S	A	Centro	741479	6945448
AP_77	1989954	15426	129	1702	1056	6931	7190	343	3	3	3	3	SM	S	S	A	Centro	741983	6946025
AP_78	269989	7297	37	494	2624	5395	4318	1454	1	1	1	1	N	N	N	B	Centro	742169	6944898
AP_79	750040	8824	85	2199	3068	8941	9200	266	2	2	0	1	MO	N	N	M	Centro	740819	6945421
AP_80	447002	4217	106	2265	2264	8142	8401	707	2	2	1	0	N	N	N	B	Centro	741257	6945360
AP_81	955016	7828	122	2974	2266	8144	8403	174	3	3	1	1	SM	N	S	M	Centro	740900	6945831
AP_82	999024	6404	156	1413	1883	7023	7282	853	3	5	3	2	N	N	N	B	Centro	742389	6945642
AP_83	449988	5844	77	662	2124	7719	4476	1357	2	1	0	1	SM	S	N	M	Centro	742378	6945091
AP_84	893582	8198	109	1855	1088	6749	7008	669	3	4	3	2	N	S	S	M	Centro	742673	6945999
AP_85	1539996	7549	204	1385	1187	7075	7334	264	4	4	2	2	SM	S	N	M	Centro	741633	6945969
AP_86	748044	7556	99	1706	1516	7391	7650	392	3	1	0	1	MO	N	S	B	Centro	741499	6945793
AP_87	314985	4565	69	1927	2406	8284	8543	880	2	1	0	0	SM	N	N	B	Centro	741487	6945279
AP_88	285032	3608	79	2236	2235	8113	8372	620	1	2	0	0	MO	N	N	B	Centro	741169	6945362
AP_89	849990	9770	87	949	1955	7843	8102	947	3	2	1	2	SM	N	S	A	Centro	741857	6945323
AP_90	2380000	11900	200	1131	1913	7801	8060	462	4	4	3	4	N	S	S	A	Centro	741866	6945847
AP_91	1259967	9921	127	2367	623	6309	6568	213	3	3	3	2	N	S	S	A	Agrônômica	742609	6946804
AP_92	780042	8966	87	2167	423	6110	6369	125	3	3	1	1	MO	N	N	A	Agrônômica	742505	6946729
AP_93	759032	9608	79	4132	2388	2604	3775	277	3	2	1	1	N	S	N	A	Agrônômica	743874	6947377
AP_94	1275030	8226	155	2618	874	5478	5737	67	4	3	1	1	N	N	N	M	Agrônômica	742504	6947027
AP_95	1690080	8048	210	2109	365	6051	6310	117	4	4	1	2	SM	S	N	M	Agrônômica	742491	6946700
AP_96	490025	5765	85	4224	2480	2694	3865	397	3	2	1	1	N	N	N	B	Agrônômica	743792	6947284
AP_97	438978	6362	69	3962	2218	2774	3945	280	2	2	1	1	SM	N	N	A	Agrônômica	743732	6947440
AP_98	650016	6771	96	4637	2893	2594	3765	249	3	2	1	1	SM	N	S	A	Agrônômica	744036	6947270
AP_99	959016	10312	93	3349	1605	5738	5997	307	3	2	1	2	MO	S	S	A	Agrônômica	743194	6947463
AP_100	976165	9665	101	3284	1540	5673	5932	176	2	3	2	1	N	S	S	A	Agrônômica	742969	6947480
AP_101	899964	7692	117	2466	722	5786	6045	147	3	2	1	1	MO	S	S	A	Agrônômica	742587	6947034
AP_102	1299987	11111	117	5283	3539	2361	3532	122	3	4	3	3	N	S	S	A	Agrônômica	744194	6947259
AP_103	2299938	12921	178	2278	534	6220	6479	116	4	3	1	2	MO	S	N	A	Agrônômica	742517	6946839
AP_104	980000	7840	125	2933	1189	5321	5580	221	3	2	1	2	N	S	S	A	Agrônômica	742753	6947267
AP_105	329976	4583	72	3920	2176	2927	4098	311	2	1	0	1	N	N	N	B	Agrônômica	743581	6947461
AP_106	769965	7333	105	3229	1485	5618	5877	284	3	2	1	1	SM	S	S	M	Agrônômica	743016	6947382

AP_107	677996	8921	76	3962	2218	2774	3945	263	3	2	1	2	MO	N	N	A	Agronômica	743737	6947457
AP_108	978027	7701	127	3227	1483	5616	5875	314	3	3	2	2	SM	N	S	A	Agronômica	742985	6947328
AP_109	780042	8966	87	2200	456	6143	6402	164	3	3	1	1	SM	N	S	A	Agronômica	742549	6946752
AP_110	1069037	9991	107	2987	1243	5376	5635	212	3	3	1	2	N	S	S	A	Agronômica	742812	6947336
AP_111	829955	7217	115	4290	2546	2446	3617	379	2	2	1	2	N	S	S	A	Agronômica	743950	6947171
AP_112	986034	9667	102	2984	1240	5372	5631	240	3	4	3	2	N	S	S	A	Agronômica	742788	6947279
AP_113	1289970	11727	110	3349	1605	5738	5997	286	3	2	1	2	SM	S	S	A	Agronômica	743182	6947480
AP_114	1900000	11875	160	2302	558	6244	6503	179	3	5	3	3	MO	S	S	A	Agronômica	742575	6946809
AP_115	550041	6627	83	4290	2546	2446	3617	369	2	2	1	2	SM	S	S	A	Agronômica	743948	6947188
AP_116	527972	6947	76	3271	1527	5660	5919	302	3	2	1	1	SM	N	N	M	Agronômica	743067	6947392
AP_117	830000	10375	80	3318	1574	5707	5966	323	2	3	2	2	SM	S	S	A	Agronômica	742920	6947278
AP_118	1390040	9145	152	2006	262	6240	6499	154	3	1	1	2	SM	S	S	A	Agronômica	742500	6946583
AP_119	693780	11190	62	2171	427	6114	6373	256	3	1	0	1	N	S	S	A	Agronômica	742613	6946613
AP_120	539981	7397	73	3891	2147	2898	4069	322	2	2	1	1	MO	N	S	M	Agronômica	743429	6947500
AP_121	1949043	13823	141	3240	1496	5629	5888	224	3	3	1	2	SM	S	S	A	Agronômica	742996	6947440
AP_122	1600029	11511	139	2869	1125	5250	5509	72	3	2	1	1	MO	N	S	A	Agronômica	742557	6947232
AP_123	415020	6917	60	4224	2480	2694	3865	376	2	1	0	1	SM	N	N	B	Agronômica	743801	6947303
AP_124	1980055	7645	259	2705	961	6025	6284	70	4	4	2	2	SM	S	S	M	Agronômica	742577	6947275
AP_125	1229984	10982	112	4096	2352	3103	4274	360	3	3	3	2	N	S	S	A	Agronômica	743441	6947456
AP_126	1150065	8519	135	2477	733	5797	6056	175	3	3	1	1	SM	S	S	A	Agronômica	742614	6947022
AP_127	1545049	13673	113	5283	3539	2361	3532	102	3	4	3	3	N	S	S	A	Agronômica	744210	6947272
AP_128	599976	5128	117	2466	722	5786	6045	158	2	2	1	1	N	S	S	B	Agronômica	742598	6947028
AP_129	2100120	7955	264	2318	574	6260	6519	140	4	4	2	2	SM	S	S	M	Agronômica	742549	6946879
AP_130	539978	7297	74	4071	2327	2795	3966	378	2	2	1	1	SM	S	S	A	Agronômica	743703	6947346
AP_131	3600090	16290	221	2032	288	6266	6525	106	3	5	3	3	SM	S	S	A	Agronômica	742452	6946593
AP_132	1800000	10000	180	3169	1425	4945	5204	91	3	5	2	2	SM	S	S	A	Agronômica	742781	6947465
AP_133	1800000	11250	160	2302	558	6244	6503	181	3	5	3	3	N	S	S	A	Agronômica	742581	6946826
AP_134	1679958	12537	134	2789	1045	5177	5436	126	3	4	1	1	SM	S	N	A	Agronômica	742661	6947306
AP_135	959000	6850	140	2043	299	5985	6244	136	3	4	1	1	N	S	S	A	Agronômica	742495	6946637
AP_136	430010	6143	70	6882	5138	1827	2105	1052	3	2	1	1	SM	N	N	M	Trindade	744142	6945971
AP_137	495040	4420	112	6536	4792	1481	1537	1031	2	3	1	1	MO	N	N	M	Trindade	744516	6945950
AP_138	489971	7313	67	5948	5882	2061	984	1974	2	1	0	1	N	N	N	M	Trindade	744877	6945053
AP_139	479988	7164	67	6778	5034	2351	1274	1314	2	2	1	1	SM	S	S	M	Trindade	744573	6945671
AP_140	536276	10313	52	6033	5852	2146	1069	1883	2	1	0	1	N	N	S	A	Trindade	744992	6945178
AP_141	870010	9158	95	6316	5605	1366	1352	1595	4	3	2	2	SM	S	S	A	Trindade	744940	6945463
AP_142	650048	5804	112	4302	5755	1992	915	1757	4	3	2	2	SM	N	S	M	Trindade	744527	6945224
AP_143	445036	4588	97	3968	6683	1952	875	1900	2	2	1	1	N	N	S	B	Trindade	744363	6945080
AP_144	495016	5756	86	6726	4982	2402	1325	1251	3	2	1	1	SM	N	N	M	Trindade	744581	6945735
AP_145	969025	11675	83	6356	5645	1326	1392	1602	3	3	1	2	MO	S	S	A	Trindade	745004	6945478
AP_146	858990	9042	95	6316	5605	1366	1352	1653	3	3	1	2	SM	S	S	A	Trindade	745022	6945430
AP_147	369990	8222	45	4956	3212	2187	2465	353	1	1	1	1	N	N	S	A	Trindade	744180	6946744
AP_148	349992	4861	72	7043	5299	2546	1469	1345	3	2	0	1	N	N	N	B	Trindade	744821	6945688
AP_149	439976	5116	86	6504	4760	1449	1727	886	2	2	1	1	SM	N	N	B	Trindade	744412	6946093
AP_150	974974	10714	91	5327	3583	1911	3085	303	3	2	1	2	N	S	S	A	Trindade	744331	6946697
AP_151	298998	6102	49	5666	3922	2233	2511	539	1	1	0	1	SM	N	N	B	Trindade	744376	6946444
AP_152	445014	6642	67	6692	4948	1781	2955	164	2	2	1	1	SM	N	N	M	Trindade	744384	6946826
AP_153	349011	6123	57	4308	5684	1921	844	1825	2	1	0	1	MO	N	N	B	Trindade	744596	6945160
AP_154	299962	3614	83	5494	3750	1649	1927	754	1	2	0	1	N	N	N	B	Trindade	744385	6946227
AP_155	590019	8551	69	6767	5023	2362	1285	1294	2	2	1	2	SM	S	S	M	Trindade	744555	6945690
AP_156	287980	4235	68	6970	5226	2643	1566	1165	3	1	0	1	MO	N	S	B	Trindade	744768	6945860
AP_157	489964	5632	87	6393	6019	1288	1429	1522	3	2	1	1	MO	N	N	M	Trindade	745041	6945579
AP_158	315000	4200	75	6970	5226	2643	1566	1194	2	1	0	1	N	N	N	B	Trindade	744813	6945843
AP_159	580000	7250	80	6215	4471	1362	2537	949	3	2	1	1	N	N	S	M	Trindade	744595	6946042
AP_160	399024	5542	72	4197	7117	2386	1309	2211	2	2	1	2	N	N	S	B	Trindade	744111	6944793
AP_161	749970	8333	90	4269	5788	2025	948	1699	3	1	1	2	MO	N	N	M	Trindade	744520	6945281

AP_162	599012	8809	68	5136	3392	2367	2494	310	2	2	1	1	MO	N	S	A	Trindade	744270	6946721
AP_163	929961	10449	89	5181	3437	2412	2539	307	3	2	1	2	SM	S	S	A	Trindade	744236	6946752
AP_164	330000	4125	80	6882	5138	1827	2105	1096	2	2	1	1	SM	N	N	B	Trindade	744110	6945935
AP_165	598980	8940	67	5407	3663	1974	2252	476	2	2	1	1	SM	N	S	A	Trindade	744367	6946509
AP_166	639975	6095	105	5971	5465	2084	1007	1605	3	3	1	2	MO	N	N	M	Trindade	744721	6945398
AP_167	689000	10600	65	5398	3654	2063	2341	388	2	2	1	2	SM	N	S	A	Trindade	744370	6946598
AP_168	457954	4673	98	3968	6683	1952	875	1898	3	2	1	1	SM	N	N	M	Trindade	744386	6945082
AP_169	549998	4661	118	4132	6557	1826	749	1926	3	3	2	1	N	N	S	M	Trindade	744470	6945053
AP_170	530024	6023	88	6944	5200	2617	1540	1212	3	2	1	1	N	N	N	M	Trindade	744749	6945806
AP_171	299988	5769	52	6970	5226	2643	1566	1137	2	1	0	1	SM	N	N	M	Trindade	744826	6945908
AP_172	690000	9200	75	5181	3437	2412	2539	263	2	2	1	1	N	S	N	A	Trindade	744241	6946810
AP_173	790023	5603	141	6341	4597	1165	1757	959	3	3	3	1	SM	N	S	M	Trindade	744580	6946030
AP_174	319995	4923	65	5494	3750	1649	1927	777	2	2	0	1	N	N	N	B	Trindade	744468	6946202
AP_175	360000	4800	75	7052	5308	2555	1478	1333	2	1	0	1	N	N	S	B	Trindade	744900	6945726
AP_176	635035	7471	85	6504	4760	1449	1727	861	3	2	1	1	MO	N	S	M	Trindade	744438	6946118
AP_177	319995	4923	65	5427	3683	1716	1994	667	3	1	0	1	N	N	N	B	Trindade	744438	6946312
AP_178	549990	7857	70	6699	4955	1644	1922	737	2	1	1	2	N	N	N	M	Trindade	744275	6946261
AP_179	589968	8194	72	6858	5114	2623	1546	1165	2	1	1	2	N	N	S	A	Trindade	744592	6945823
AP_180	294976	4609	64	5715	3971	1986	2264	937	2	1	0	1	N	N	N	B	Trindade	744045	6946131
AP_181	350025	4667	75	7052	5308	2555	1478	1275	3	1	0	1	N	N	N	B	Trindade	744913	6945793
AP_182	479960	3692	130	5718	5558	1831	754	1728	3	3	1	1	SM	N	N	B	Trindade	744628	6945261
AP_183	340032	5152	66	5781	5567	1894	817	1743	2	1	0	1	N	N	N	B	Trindade	744721	6945258
AP_184	348000	7250	48	4120	7040	2309	1232	2115	1	1	0	1	SM	N	S	M	Trindade	744148	6944884
AP_185	440016	4944	89	3917	6837	2106	1029	1942	2	2	1	1	N	N	N	M	Trindade	744244	6945047
AP_186	505036	4244	119	3904	6746	2015	938	1902	3	3	1	2	N	N	N	B	Trindade	744295	6945082
AP_187	519979	7123	73	6033	5852	2146	1069	1903	2	2	1	1	N	N	S	M	Trindade	744999	6945159
AP_188	260000	6500	40	5446	3702	2013	2291	430	1	1	0	1	SM	N	S	B	Trindade	744373	6946554
AP_189	290004	4394	66	5665	3921	1936	2214	912	3	1	0	1	N	N	N	B	Trindade	744057	6946153
AP_190	694944	4572	152	6276	4532	1221	1797	907	2	2	1	1	N	N	S	M	Trindade	744536	6946076

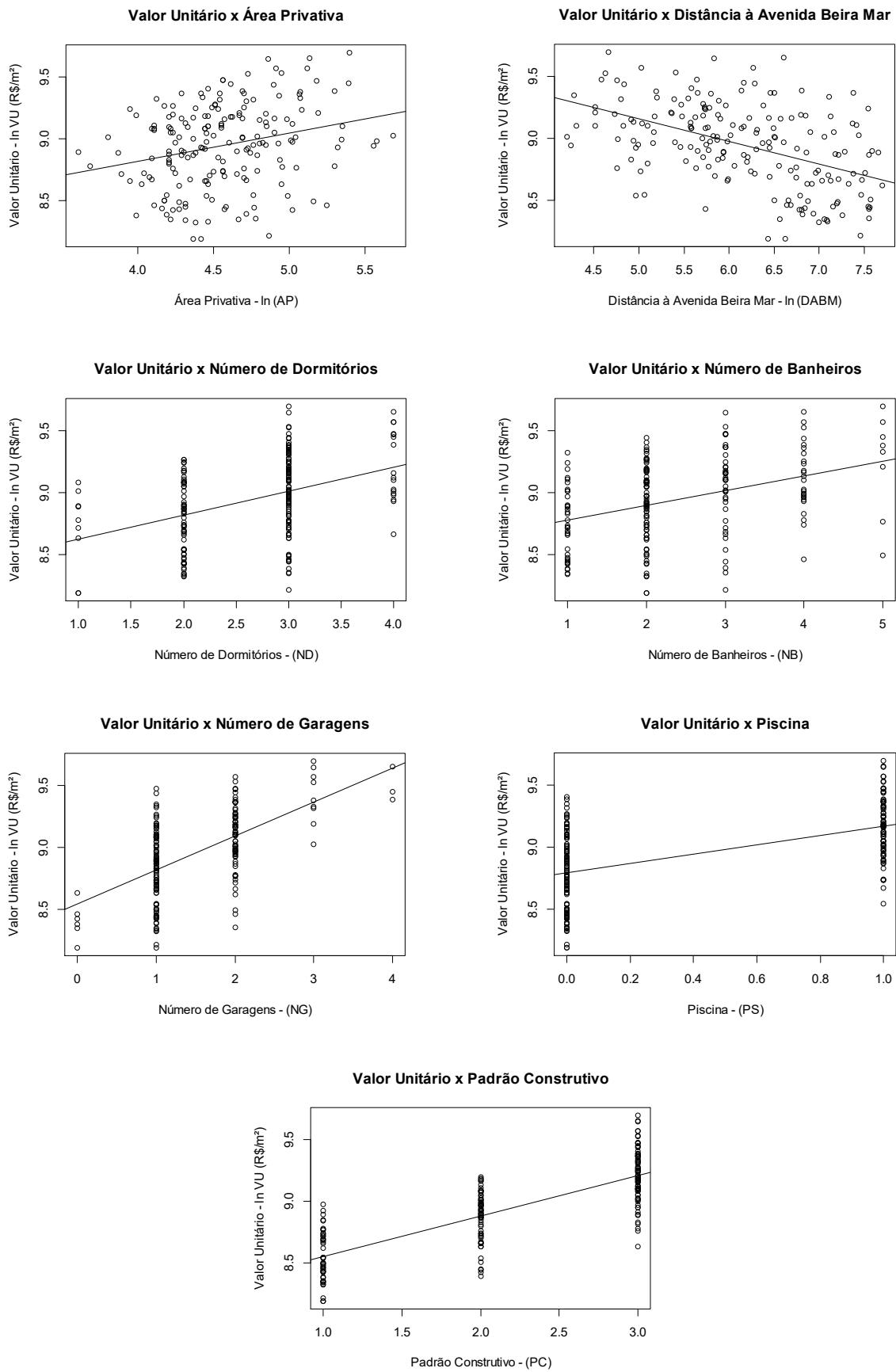
**LEGENDA:** VT → Valor Total em R\$; VU → Valor Unitário em R\$/m<sup>2</sup>; AP → Área Privativa em m<sup>2</sup>; DPXV → Distância à Praça XV de Novembro em m; DSBM → Distância ao Shopping Beira Mar em m; DSIG → Distância ao Shopping Iguatemi em m; DCTC → Distância ao Centro Tecnológico na UFSC em m; DABM → Distância à Avenida Beira Mar Norte em m; ND → Número de Dormitórios em unidades; NB → Número de Banheiros Sociais em unidades; NS → Número de Suítes em unidades; NG → Número de Vagas de Garagem em unidades; MO → Existência de Mobília; PS → Existência de Piscina no Condomínio; CH → Churrasqueira; PC → Padrão Construtivo; COORD\_E → Coordenada UTM E; COORD\_N → Coordenada UTM N.

## APÊNDICE 02: AMOSTRA DE DADOS DE MERCADO (VALIDAÇÃO)

**LEGENDA:** VT → Valor Total em R\$; VU → Valor Unitário em R\$/m<sup>2</sup>; AP → Área Privativa em m<sup>2</sup>; DPXV → Distância à Praça XV de Novembro em m; DSBM → Distância ao Shopping Beira Mar em m; DSIG → Distância ao Shopping Iguatemi em m; DCTC → Distância ao Centro Tecnológico na UFSC em m; DABM → Distância à Avenida Beira Mar Norte em m; ND → Número de Dormitórios em unidades; NB → Número de Banheiros Sociais em unidades; NS → Número de Suítes em unidades; NG → Número de Vagas de Garagem em unidades; MO → Existência de Mobília; PS → Existência de Piscina no Condomínio; CH → Churrasqueira; PC → Padrão Construtivo; COORD\_E → Coordenada UTM E; COORD\_N → Coordenada UTM N.

NOME	VT	VU	AP	DPXV	DSBM	DSIG	DCTC	DABM	ND	NB	NS	NG	MO	PS	CH	PC	BRO	COORD_E	COORD_N
AV_01	1299980	11818	110	1836	1050	6937	7196	668	3	2	1	2	SM	N	N	A	Centro	742431	6945863
AV_02	535040	5632	95	2293	2292	8170	8429	720	2	3	1	0	MO	N	S	B	Centro	741257	6945309
AV_03	879954	8627	102	1999	181	6159	6418	113	3	2	1	1	SM	N	S	A	Centro	742421	6946540
AV_04	849968	7589	112	1511	1285	7160	7419	598	3	3	1	1	MO	S	S	A	Centro	742170	6945826
AV_05	457878	6834	67	818	1976	7572	4808	1060	2	3	2	1	N	N	N	M	Centro	742254	6945360
AV_06	441000	7350	60	1198	1271	7159	7418	379	2	1	0	0	MO	N	S	M	Centro	741620	6945842
AV_07	748974	8709	86	1767	981	6868	7127	583	1	2	1	1	SM	S	S	A	Centro	742345	6945919
AV_08	849873	7521	113	1356	1480	7433	7692	644	3	3	1	2	SM	N	N	B	Centro	742032	6945719
AV_09	1350079	7459	181	1848	1062	6949	7208	478	4	1	1	2	SM	N	N	A	Centro	742308	6946017
AV_10	1290000	10000	129	2499	1791	7669	7928	99	3	3	1	2	MO	N	N	M	Centro	740816	6945883
AV_11	779960	5270	148	2092	618	6505	6764	210	3	2	1	1	N	N	N	M	Centro	742293	6946303
AV_12	1180015	10261	115	3920	1855	7733	7992	389	3	3	1	1	MO	S	S	A	Centro	741100	6945670
AV_13	1049997	9633	109	1210	1992	7880	8139	604	3	3	1	2	SM	N	N	M	Centro	741968	6945735
AV_14	1400000	17500	80	2381	2380	8258	8517	726	3	3	3	2	N	S	S	A	Centro	741501	6945448
AV_15	770032	11324	68	3111	2404	8282	8541	394	2	2	1	1	N	N	N	M	Centro	740953	6945523
AV_16	2799930	14070	199	1752	1045	6920	7179	100	3	4	3	3	MO	S	S	M	Centro	741540	6946110
AV_17	889900	8090	110	1958	606	6493	6752	349	3	3	1	2	SM	N	N	M	Centro	742383	6946193
AV_18	613149	10757	57	2156	412	6099	6358	261	2	1	0	1	N	S	N	A	Agronômica	742617	6946629
AV_19	799000	8500	94	4572	2828	2529	3700	192	3	2	1	2	N	N	S	M	Agronômica	744074	6947309
AV_20	689986	8734	79	2933	1189	5321	5580	198	2	3	1	2	SM	S	S	M	Agronômica	742739	6947285
AV_21	1136956	10726	106	2987	1243	5376	5635	202	3	2	1	2	N	S	S	A	Agronômica	742812	6947350
AV_22	950053	8879	107	4209	2465	3216	4387	407	3	3	3	2	N	S	S	M	Agronômica	743518	6947383
AV_23	2300100	10455	220	2109	365	6051	6310	124	4	4	2	2	SM	N	N	A	Agronômica	742488	6946684
AV_24	375013	4747	79	5807	5593	1920	843	1775	2	1	0	1	N	N	S	B	Trindade	744742	6945231
AV_25	680021	5191	131	6795	5051	2333	1256	1294	2	2	1	2	MO	N	N	M	Trindade	744589	6945695
AV_26	679952	6071	112	4302	5755	1992	915	1771	3	2	2	2	SM	N	S	M	Trindade	744521	6945211
AV_27	684886	9382	73	6621	4877	1566	1844	812	2	2	1	1	MO	N	S	A	Trindade	744289	6946183
AV_28	329976	4583	72	6882	5138	1827	2105	1079	2	2	0	1	N	N	N	B	Trindade	744125	6945949
AV_29	989985	10645	93	5181	3437	2412	2539	255	3	2	1	2	SM	S	S	A	Trindade	744262	6946801
AV_30	350035	4795	73	7043	5299	2546	1469	1361	3	2	0	1	N	N	N	B	Trindade	744825	6945674
AV_31	619990	7294	88	6504	4760	1449	1727	902	3	2	1	1	SM	N	S	A	Trindade	744400	6946079
AV_32	875022	10671	82	6316	5605	1366	1352	1630	2	2	1	2	SM	S	S	A	Trindade	745021	6945457
AV_33	870010	9158	95	6338	5627	1385	1374	1647	4	3	2	2	N	S	N	A	Trindade	744960	6945417
AV_34	542997	8619	63	6725	4981	1670	1948	718	2	1	1	1	N	N	S	M	Trindade	744290	6946279
AV_35	650040	5417	120	5106	6521	1790	713	1951	3	3	2	1	SM	N	S	B	Trindade	744472	6945030

### APÊNDICE 03: GRÁFICOS DE DISPERSÃO E LINEARIDADE



## APÊNDICE 04: ESTATÍSTICAS DE 50 REGRESSÕES LOCAIS DE GWR

N	EST INTER	SE INTER	T INTER	EST AP	SE AP	T AP	EST DABM	SE DABM	T DABM	EST NG	SE NG	T NG	EST PC	SE PC	T PC	Y HAT	RESI DUO	LOCAL R <sup>2</sup>	INFLU ENCE	DIST COOK
1	11,41	0,70	16,32	-0,61	0,15	-3,96	-0,11	0,05	-2,39	0,23	0,05	4,67	0,24	0,05	5,26	9,09	0,07	0,88	0,30	0,00
2	11,41	0,71	16,08	-0,59	0,17	-3,56	-0,11	0,05	-2,34	0,23	0,05	4,69	0,23	0,05	5,12	8,84	-0,08	0,87	0,47	0,01
3	11,94	0,66	18,19	-0,71	0,16	-4,34	-0,12	0,04	-2,77	0,20	0,04	4,65	0,20	0,04	4,72	8,85	0,07	0,85	0,48	0,01
4	12,06	0,69	17,47	-0,82	0,16	-5,09	-0,09	0,04	-2,16	0,17	0,05	3,42	0,15	0,04	3,45	8,98	0,11	0,86	0,20	0,00
5	10,74	0,59	18,35	-0,49	0,12	-4,16	-0,09	0,04	-2,37	0,19	0,04	4,42	0,22	0,03	6,23	8,93	0,03	0,89	0,13	0,00
6	10,47	0,68	15,33	-0,38	0,13	-2,84	-0,12	0,04	-2,59	0,17	0,05	3,63	0,18	0,04	4,64	8,92	0,16	0,89	0,19	0,01
7	11,07	0,63	17,44	-0,51	0,12	-4,30	-0,12	0,04	-2,88	0,26	0,05	5,55	0,19	0,04	4,54	9,42	0,06	0,89	0,33	0,00
8	11,63	0,61	18,96	-0,66	0,14	-4,56	-0,12	0,04	-3,23	0,22	0,04	5,34	0,24	0,04	6,86	9,05	-0,19	0,90	0,10	0,00
9	11,70	0,58	20,21	-0,65	0,14	-4,54	-0,12	0,04	-3,32	0,21	0,04	5,31	0,22	0,04	5,94	9,12	-0,10	0,87	0,27	0,00
10	10,41	0,70	14,81	-0,35	0,13	-2,76	-0,11	0,04	-2,71	0,21	0,05	3,89	0,16	0,04	4,06	9,48	0,09	0,89	0,35	0,01
11	10,86	0,63	17,18	-0,46	0,12	-3,98	-0,11	0,04	-2,81	0,25	0,05	5,21	0,18	0,04	4,41	8,90	-0,07	0,89	0,22	0,00
12	10,81	0,56	19,15	-0,55	0,12	-4,69	-0,08	0,04	-2,03	0,18	0,04	4,19	0,22	0,03	6,78	9,08	-0,07	0,90	0,11	0,00
13	11,42	0,74	15,40	-0,61	0,15	-4,02	-0,12	0,05	-2,57	0,26	0,05	4,90	0,23	0,05	4,88	9,26	0,00	0,88	0,54	0,00
14	12,07	0,73	16,55	-0,85	0,17	-5,08	-0,08	0,05	-1,75	0,17	0,05	3,18	0,15	0,05	3,16	9,26	0,10	0,86	0,45	0,01
15	11,40	0,70	16,26	-0,61	0,15	-4,02	-0,11	0,05	-2,41	0,23	0,05	4,69	0,24	0,05	5,26	9,14	0,00	0,88	0,43	0,00
16	12,07	0,68	17,74	-0,82	0,16	-5,09	-0,10	0,04	-2,28	0,18	0,05	3,49	0,15	0,04	3,56	9,19	-0,19	0,86	0,24	0,01
17	10,62	0,64	16,68	-0,42	0,12	-3,41	-0,11	0,04	-2,66	0,20	0,05	4,27	0,19	0,04	4,97	8,95	0,12	0,89	0,18	0,00
18	12,33	0,79	15,66	-0,88	0,18	-4,92	-0,09	0,05	-1,86	0,18	0,06	3,18	0,15	0,05	3,02	9,43	0,05	0,86	0,43	0,00
19	12,01	0,69	17,33	-0,81	0,16	-5,03	-0,10	0,04	-2,27	0,17	0,05	3,41	0,15	0,05	3,36	9,00	0,02	0,87	0,64	0,00
20	11,61	0,62	18,77	-0,68	0,14	-4,73	-0,11	0,04	-3,00	0,21	0,04	5,15	0,24	0,03	7,20	9,32	0,13	0,90	0,19	0,00
21	11,98	0,66	18,04	-0,79	0,16	-5,04	-0,10	0,04	-2,53	0,17	0,05	3,56	0,15	0,04	3,65	8,95	0,16	0,87	0,18	0,01
22	10,57	0,66	15,95	-0,39	0,12	-3,22	-0,11	0,04	-2,78	0,22	0,05	4,29	0,17	0,04	4,33	8,93	-0,12	0,89	0,20	0,00
23	11,70	0,59	19,88	-0,65	0,15	-4,42	-0,12	0,04	-3,25	0,21	0,04	5,25	0,22	0,04	5,87	9,16	0,08	0,87	0,30	0,00
24	12,10	0,69	17,62	-0,82	0,16	-5,09	-0,10	0,04	-2,25	0,18	0,05	3,48	0,15	0,04	3,54	9,00	0,09	0,86	0,21	0,00
25	10,45	0,78	13,40	-0,34	0,14	-2,40	-0,13	0,05	-2,75	0,19	0,06	3,36	0,16	0,04	3,86	8,90	0,07	0,89	0,37	0,00
26	12,07	0,74	16,42	-0,82	0,17	-4,89	-0,10	0,05	-2,16	0,17	0,05	3,27	0,16	0,05	3,13	8,76	0,01	0,87	0,49	0,00
27	11,34	0,57	19,96	-0,67	0,12	-5,35	-0,09	0,04	-2,38	0,20	0,04	4,59	0,24	0,03	7,53	9,15	0,01	0,90	0,22	0,00
28	11,92	0,59	20,38	-0,74	0,14	-5,11	-0,12	0,04	-3,09	0,18	0,04	4,34	0,18	0,04	4,79	9,01	-0,02	0,87	0,30	0,00
29	11,36	0,69	16,39	-0,57	0,13	-4,24	-0,13	0,04	-2,96	0,29	0,05	5,76	0,20	0,04	4,55	9,20	-0,10	0,87	0,45	0,01
30	12,14	0,69	17,48	-0,83	0,16	-5,08	-0,10	0,04	-2,23	0,18	0,05	3,47	0,15	0,04	3,51	9,11	-0,19	0,85	0,20	0,01
31	12,26	0,73	16,80	-0,85	0,17	-5,05	-0,10	0,05	-2,13	0,18	0,05	3,41	0,15	0,04	3,34	9,53	0,04	0,85	0,47	0,00
32	11,77	0,65	18,01	-0,70	0,16	-4,39	-0,13	0,04	-3,19	0,17	0,05	3,61	0,18	0,04	4,45	8,75	-0,06	0,91	0,29	0,00
33	11,93	0,61	19,51	-0,76	0,15	-5,05	-0,11	0,04	-2,96	0,18	0,04	3,95	0,17	0,04	4,36	8,93	0,04	0,87	0,17	0,00
34	12,03	0,69	17,45	-0,82	0,16	-5,09	-0,10	0,04	-2,16	0,17	0,05	3,41	0,15	0,04	3,42	9,38	-0,18	0,86	0,47	0,04
35	11,60	0,61	18,87	-0,69	0,14	-4,81	-0,11	0,04	-2,94	0,21	0,04	5,09	0,25	0,03	7,28	8,93	-0,02	0,90	0,31	0,00
36	12,00	0,63	19,06	-0,78	0,15	-5,09	-0,11	0,04	-2,77	0,18	0,05	3,83	0,16	0,04	4,10	8,62	0,07	0,86	0,18	0,00
37	11,63	0,60	19,27	-0,63	0,15	-4,23	-0,12	0,04	-3,17	0,22	0,04	5,20	0,23	0,04	6,01	9,24	-0,05	0,87	0,30	0,00
38	10,51	0,75	14,08	-0,36	0,14	-2,58	-0,13	0,05	-2,76	0,18	0,05	3,44	0,16	0,04	4,10	9,39	-0,02	0,89	0,28	0,00
39	12,02	0,78	15,47	-0,73	0,19	-3,78	-0,16	0,05	-3,11	0,16	0,05	3,04	0,17	0,05	3,39	8,73	0,00	0,91	0,58	0,00
40	11,20	0,64	17,50	-0,55	0,12	-4,47	-0,12	0,04	-2,86	0,26	0,05	5,41	0,20	0,04	4,96	9,33	0,08	0,89	0,26	0,00
41	11,19	0,56	20,06	-0,60	0,11	-5,21	-0,10	0,04	-2,57	0,21	0,04	4,77	0,24	0,03	7,28	9,52	-0,07	0,89	0,32	0,00
42	10,49	0,72	14,55	-0,36	0,14	-2,65	-0,13	0,05	-2,68	0,18	0,05	3,44	0,17	0,04	4,30	9,30	0,07	0,89	0,18	0,00
43	11,53	0,60	19,32	-0,64	0,14	-4,53	-0,12	0,04	-3,20	0,20	0,04	5,33	0,22	0,04	6,13	9,18	0,09	0,92	0,45	0,01
44	10,72	0,64	16,69	-0,43	0,12	-3,67	-0,11	0,04	-2,76	0,24	0,05	4,90	0,17	0,04	4,31	8,75	-0,08	0,89	0,47	0,01
45	11,20	0,55	20,21	-0,60	0,11	-5,27	-0,10	0,04	-2,55	0,21	0,04	4,74	0,24	0,03	7,40	9,04	0,12	0,89	0,16	0,00
46	11,10	0,55	20,04	-0,61	0,12	-5,25	-0,08	0,04	-2,23	0,19	0,04	4,43	0,24	0,03	7,35	8,96	-0,04	0,90	0,10	0,00
47	10,78	0,62	17,50	-0,46	0,12	-3,88	-0,11	0,04	-2,76	0,22	0,05	4,73	0,19	0,04	4,99	9,11	-0,03	0,89	0,57	0,00
48	11,85	0,62	19,26	-0,73	0,15	-4,88	-0,12	0,04	-3,08	0,18	0,05	3,89	0,17	0,04	4,48	8,77	0,08	0,89	0,35	0,00
49	11,32	0,62	18,24	-0,70	0,14	-4,95	-0,07	0,04	-1,76	0,20	0,05	4,31	0,24	0,04	6,88	9,58	0,08	0,92	0,40	0,00
50	11,24	0,60	18,73	-0,59	0,12	-4,84	-0,11	0,04	-2,61	0,22	0,05	4,89	0,24	0,04	6,38	9,49	0,15	0,89	0,27	0,01

**LEGENDA:** (EST) → Estimativa, (SE) → Erro Padrão, (T) → Estatística t, (Y HAT) → Valores Preditos, (LOC) → Local, (DIST COOK) → Distâncias de Cook.

## APÊNDICE 05: ALGUNS CÓDIGOS UTILIZADOS NESSE ESTUDO

### Mapa Variográfico e Semivariogramas Experimental e Teóricos

---

```

dados = read.csv (file = "Invugwr.csv", sep = ";", header = T)
library(gstat)
library(raster)
library(rstudioapi)
library(sp)
library(hydroGOF)

dados<-dados[,c(1,2,3)]
names(dados)<-c("x", "y", "z")
sp::coordinates(dados) = ~ x + y
Invugwr<-dados$z
print(max(dist(dados@coords))/2)
print(min(dist(dados@coords)))

#semivariograma gstat automatico#

v = variogram(Invugwr ~ x + y, dados)
fit.variogram(v, vgm("Sph"))
fit.variogram(v, vgm("Exp"))
fit.variogram(v, vgm("Gau"))

#semivariograma experimental com formula#

v=gstat(id="Invugwr", formula = Invugwr ~ x + y, data=dados)

var_map = gstat::variogram(v, cutoff = 2178.96, width = 200, map=T)
plot(var_map, main="Mapa Variográfico de (VU) - GWR")

var_exp = gstat::variogram(v, cutoff = 2178.96, width = 200)
plot(var_exp, main="Semivariograma Experimental de ln (VU) - GWR", xlab="Distância (m)",
     ylab="Semivariância de (VU)", cex=1, pch=19)

#semivariogramas teóricos#

fit.sph = fit.variogram(var_exp, vgm(286183.748, "Sph", 1709.865, 1041.479)) #contribuição, modelo, alcance, pepita
plot(var_exp, fit.sph,
     main="Semivariograma Teórico Esférico de GWR", xlab="Distância (m)",
     ylab="Semivariância de (VU) - GWR", cex=1, pch=19)

fit.exp = fit.variogram(var_exp, vgm(534291.4, "Exp", 1919.145, 0.0))
plot(var_exp, fit.exp,
     main="Semivariograma Teórico Exponencial de GWR", xlab="Distância (m)",
     ylab="Semivariância de (VU) - GWR", cex=1, pch=19)

fit.gau = fit.variogram(var_exp, vgm(214143.53, "Gau", 569.7336, 18589.45))
plot(var_exp, fit.gau,
     main="Semivariograma Teórico Gaussiano de GWR", xlab="Distância (m)",
     ylab="Semivariância de (VU) - GWR", cex=1, pch=19)

#validacao cruzada#

xvalid.sph = krige.cv(z ~ x + y, locations = dados, model = fit.sph)
plot(xvalid.sph$var1.pred ~ Invugwr, cex = 1, lwd = 2)
abline(0, 1, col = "lightgrey", lwd = 2)
lm_sph = lm(xvalid.sph$var1.pred ~ Invugwr)
abline(lm_sph, col = "red", lwd = 2)
r2_sph = summary(lm_sph)$r.squared
rmse_sph = hydroGOF::rmse(xvalid.sph$var1.pred, Invugwr)

```

```

xvalid.exp = krig.cv(z ~ x + y, locations = dados, model = fit.exp)
plot(xvalid.exp$var1.pred ~ Invugwr, cex = 1, lwd = 2)
abline(0, 1, col = "lightgrey", lwd = 2)
lm_exp = lm(xvalid.exp$var1.pred ~ Invugwr)
abline(lm_exp, col = "red", lwd = 2)
r2_exp = summary(lm_exp)$r.squared
rmse_exp = hydroGOF::rmse(xvalid.exp$var1.pred, Invugwr)

xvalid.gau = krig.cv(z ~ x + y, locations = dados, model = fit.gau)
plot(xvalid.gau$var1.pred ~ Invugwr, cex = 1, lwd = 2)
abline(0, 1, col = "lightgrey", lwd = 2)
lm_gau = lm(xvalid.gau$var1.pred ~ Invugwr)
abline(lm_gau, col = "red", lwd = 2)
r2_gau = summary(lm_gau)$r.squared
rmse_gau = hydroGOF::rmse(xvalid.gau$var1.pred, Invugwr)

#mostrar resultados#
df.r2 = data.frame(r2_exp, r2_gau, r2_sph)
df.rmse = data.frame(rmse_exp, rmse_gau, rmse_sph)
temp = data.frame(cbind(t(df.r2), t(df.rmse)))
colnames(temp) = c('R2', 'RMSE')
rnames = gsub('r2_',' ', rownames(temp))
rownames(temp) = rnames
print(temp)

```

## Raster de Busca de Vizinho + Próximo

---

```

library(raster)
subjectID <- c("AV_01", ..., "AV_34")
subxy <- matrix(c(742431,6945863, ..., 744472,6945030), ncol=2, byrow=TRUE)

blockID <- c("AP_01", ..., "AP_180")
blockxy <- matrix(c(742452,6946217, ..., 744536,6946076), ncol=2, byrow=TRUE)

d <- pointDistance(subxy, blockxy, lonlat=FALSE)
d
r <- apply(d, 1, which.min)
r
p <- data.frame(subject=subjectID, block=blockID[r])
p
plot(rbind(blockxy, subxy), xlab='Longitude', ylab='Latitude')
points(blockxy, col="red", pch=20, cex=2)
points(subxy, col="blue", pch=20, cex=2)
text(subxy, subjectID, pos=1)
text(blockxy, blockID, pos=1)
for (i in 1:nrow(subxy)) {arrows(subxy[i,1], subxy[i,2], blockxy[r[i],1], blockxy[r[i],2])}

```

## APÊNDICE 06: MODELAGEM CLÁSSICA (SEM AS VARIÁVEIS DE BAIRRO)

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data Set : Tabela SB  
 Dependent Variable : lnVU Number of Observations: 188  
 Mean dependent var : 8,947460 Number of Variables : 8  
 S.D. dependent var : 0,333141 Degrees of Freedom : 180

R-squared : 0,7958740 F-statistic : 100,258  
 Adjusted R-squared : 0,7879360 Prob(F-statistic) : 000,000  
 Sum squared residual: 4,2590500 Log likelihood : 89,2547  
 Sigma-square : 0,0236614 Akaike info criterion : -162,509  
 S.E. of regression : 0,1538230 Schwarz criterion : -136,618  
 Sigma-square ML : 0,0226545  
 S.E.of regression ML: 0,1505140

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	10,436800	0,270668	38,5592	0,00000
ln (AP)	-0,412255	0,058423	-7,05631	0,00000
ln (DABM)	-0,102764	0,015366	-6,68776	0,00000
ND	0,066858	0,022846	2,92649	0,00387
NB	0,044783	0,018016	2,48571	0,01384
NG	0,175033	0,022513	7,77449	0,00000
PS	0,097637	0,027879	3,50208	0,00058
PC	0,200163	0,017918	11,1709	0,00000

### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 89,32

### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4,9546	0,08397

### DIAGNOSTICS FOR HETEROSKEDASTICITY

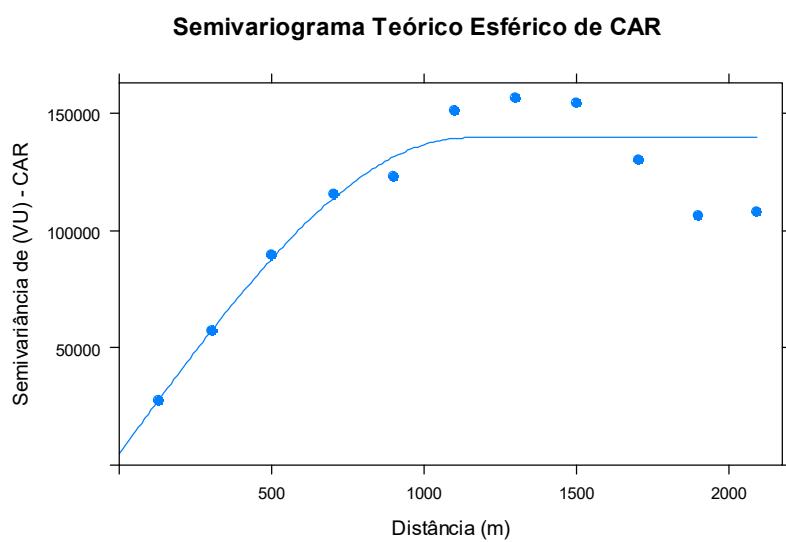
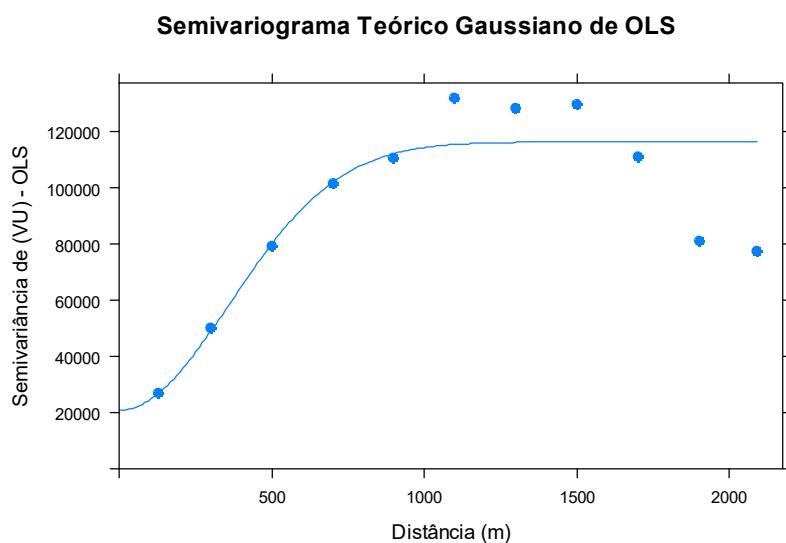
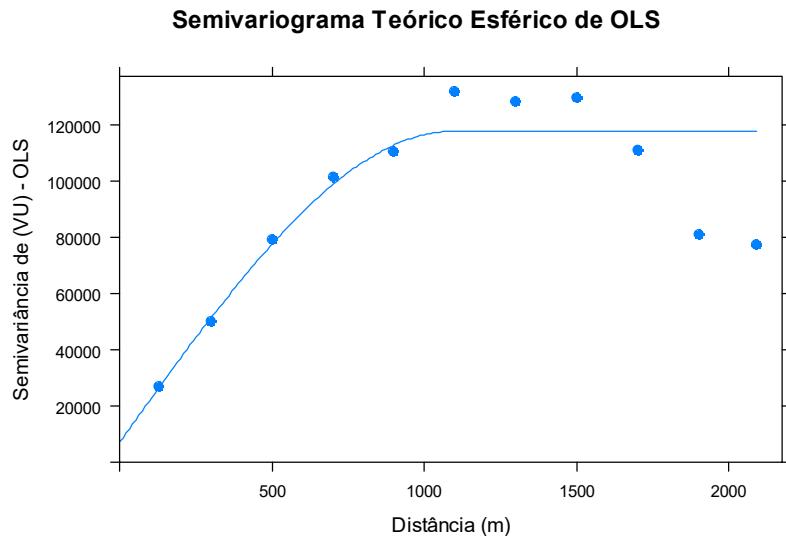
#### RANDOM COEFFICIENTS

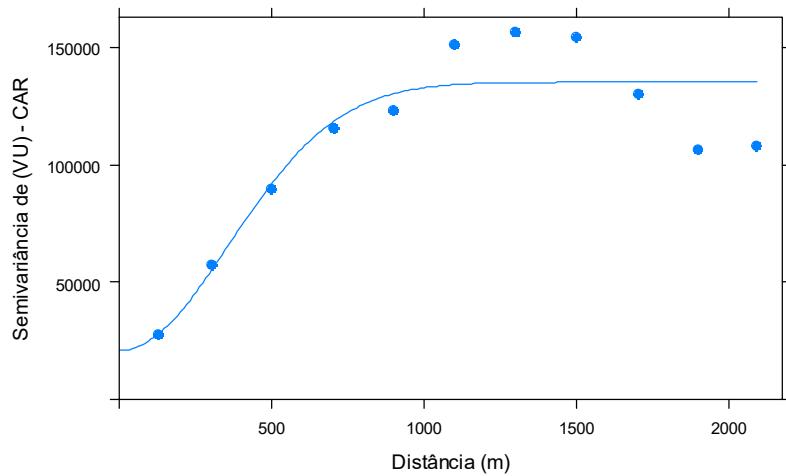
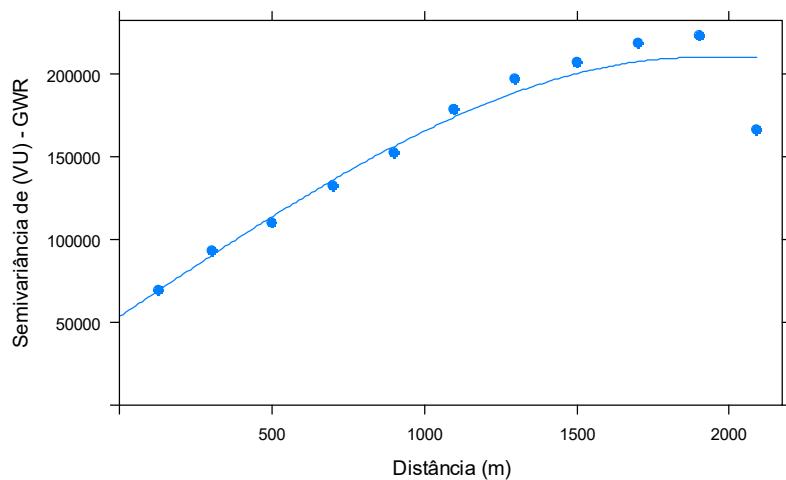
TEST	DF	VALUE	PROB
Breusch-Pagan test	7	5,3832	0,61331
Koenker-Bassett test	7	8,6434	0,27928

### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : 525

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0,1313	6,9834	0,00000
Lagrange Multiplier (lag)	1	7,3667	0,00664
Robust LM (lag)	1	0,1318	0,71662
Lagrange Multiplier (error)	1	16,3486	0,00005
Robust LM (error)	1	9,1136	0,00254
Lagrange Multiplier (SARMA)	2	16,4803	0,00026

**APÊNDICE 07: SEMIVARIOGRAMAS TEÓRICOS DAS MODELAGENS**

**Semivariograma Teórico Gaussiano de CAR****Semivariograma Teórico Esférico de GWR****Semivariograma Teórico Gaussiano de GWR**