

**ESTUDO DA INFLUÊNCIA DA
ACESSIBILIDADE NO VALOR DE LOTES
URBANOS ATRAVÉS DO USO DE REDES
NEURAIAS**

NAIR CRISTINA MARGARIDO BRONDINO

Tese apresentada à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Doutor em Engenharia Civil - Transportes.

ORIENTADOR: Prof. Assoc. Antônio Néilson
Rodrigues da Silva

São Carlos
1999

DEDICATÓRIA

*Ao meu marido, Odney, e aos meus filhos, Bárbara e
Brunno, pela compreensão e solidariedade, pelo amor
sem cobranças e principalmente por sempre terem
entendido os meus momentos de ausência.*

*Aos meus pais, Néelson e Zuleica, que nunca mediram
sacrifícios para me dar uma boa educação.*

*Aos meus irmãos e a aqueles familiares e amigos, que sempre
torceram por mim.*

AGRADECIMENTOS

À FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, pela bolsa de estudos e todo o apoio financeiro concedido.

Ao meu orientador, Prof. Associado Antônio Néelson Rodrigues da Silva, não só pela orientação mas também pela atenção, compreensão e solidariedade dispensadas nos momentos mais difíceis da realização deste trabalho.

Ao Departamento de Transportes, pelo financiamento e intervenção junto à Prefeitura de São Carlos, por ocasião da aquisição das cópias das fichas do cadastro da cidade.

Ao Prof. Dr. André C. Ponce de Leon Ferreira Camargo e ao assessor externo da FAPESP, cujas sugestões serviram, em muito, para o aprimoramento deste trabalho.

Ao meu marido, Odney Carlos Brondino, pela ajuda, especialmente nas etapas de levantamento bibliográfico e correção.

À minha prima, colega de departamento e amiga, Sandra A. Margarido Bertollo, por estar sempre disposta a ajudar no que for preciso e pelo apoio moral.

Aos demais colegas do Departamento de Transportes e em especial: Adelaide Figueiredo, Sandra Rodgher, Renato Lima, Archimedes A. Raia Jr., Eliane Viviani, Vânia Pampolha, Jônatas Melo e Elias Matsuo.

Aos funcionários do Departamento de Transportes e em especial: Lilian, Heloísa, Angélica, Magali, Sueli, Carlos e Fio.

Enfim, agradeço a todos aqueles que, de uma forma ou de outra, contribuíram para a realização deste trabalho.

SUMÁRIO

1. INTRODUÇÃO	1
1.1. OBJETIVOS.....	3
1.2. MÉTODO.....	4
1.2. ESTRUTURA DO TRABALHO	5
2. AVALIAÇÃO DE IMÓVEIS	5
2.1. ALGUMAS CONSIDERAÇÕES SOBRE ACESSIBILIDADE E O VALOR DO SOLO.....	6
2.2. SITUAÇÃO ATUAL DA AVALIAÇÃO DE IMÓVEIS	9
2.3. AVALIAÇÃO COLETIVA DE IMÓVEIS	12
2.3.1. COLETA DE DADOS	14
2.3.2. METODOLOGIA APLICADA NA CONSTRUÇÃO DE UMA PGV	15
2.3.2.1. MODELOS PREDETERMINADOS.....	16
2.3.2.2. MODELOS PROBABILÍSTICOS OU INFERENCIAIS.....	18
2.4. ALGUNS MODELOS DE AVALIAÇÃO NACIONAIS	19
2.5. EXEMPLOS DE MODELOS DE AVALIAÇÃO DE IMÓVEIS ENCONTRADOS NA LITERATURA INTERNACIONAL.....	23
2.6. PRÓS E CONTRAS DOS IMPOSTOS SOBRE A PROPRIEDADE	28
3. REDES NEURAIS, REGRESSÃO LINEAR MÚLTIPLA E SIG	30
3.1 ASPECTOS DA INTELIGÊNCIA ARTIFICIAL	30
3.1.1 ANALOGIA BIOLÓGICA	31
3.1.2. HISTÓRICO – REDES NEURAIS	32
3.1.3. REDES NEURAIS ARTIFICIAIS	33
3.1.3.1. O NEURÔNIO ARTIFICIAL	35
3.1.3.2. TIPOS DE FUNÇÃO DE ATIVAÇÃO.....	36
3.1.3.3. TOPOLOGIA DA REDE.....	37
3.1.3.4. ALGORITMOS DE APRENDIZADO	37
3.1.3.5. A REDE <i>Multlayer Perceptron</i>	38
3.1.3.5.1. ALGORITMO DE APRENDIZADO PARA UMA REDE MLP	40
3.2. ANÁLISE DE REGRESSÃO	44
3.2.1. O MODELO LINEAR	44
3.2.1.1. SUPOSIÇÕES DO MODELO DE REGRESSÃO LINEAR CLÁSSICO.....	45

3.2.1.1.1. LINEARIDADE DO MODELO DE REGRESSÃO	45
3.2.1.1.2. POSTO COMPLETO	46
3.2.1.1.3 VALOR ESPERADO DO DISTÚRBO	46
3.2.1.1.4. DISTÚRBIOS ESFÉRICOS	47
3.2.1.1.5. REGRESSORES NÃO-ESTOCÁSTICOS.....	48
3.2.1.1.6. NORMALIDADE.....	48
3.2.1.2. O MÉTODO DOS MÍNIMOS QUADRADOS	48
3.2.1.2.1. O VETOR DOS COEFICIENTES OBTIDO POR MÍNIMOS QUADRADOS.....	50
3.2.1.2.1.1. TESTES DE HIPÓTESE E INTERVALOS DE CONFIANÇA PARA O COEFICIENTE β_k ..	50
3.2.1.3. VERIFICAÇÃO DA ADEQUABILIDADE DA REGRESSÃO AJUSTADA	52
3.2.1.4. MULTICOLINEARIDADE.....	54
3.2.1.4.1. POSSÍVEIS SINTOMAS E ALGUMAS MEDIDAS DE MULTICOLINEARIDADE.....	54
3.2.1.4.2. POSSÍVEIS SOLUÇÕES PARA O PROBLEMA DA MULTICOLINEARIDADE.....	57
3.2.1.5. HETEROCEDASTICIDADE	58
3.2.1.6. OS EFEITOS DE INTERAÇÃO	59
3.3. OS SISTEMAS DE INFORMAÇÃO GEOGRÁFICA (SIG)	60
3.3.1. HISTÓRICO.....	62
3.3.2. TRATAMENTO DA INFORMAÇÃO.....	64
3.3.2.1. TRATAMENTO DE DADOS ALFANUMÉRICOS	64
3.3.2.2. TRATAMENTO DOS DADOS GRÁFICOS	65
3.3.3. ESTRUTURAS DE REPRESENTAÇÃO DO ESPAÇO GEOGRÁFICO.....	66
3.3.4. DIFERENÇAS ENTRE OS SIG's E O CAD	68
3.3.5. SITUAÇÃO ATUAL DO USO DE SIG NA AVALIAÇÃO IMOBILIÁRIA	69
3.3.6. ESTUDOS PRELIMINARES SOBRE ESTATÍSTICA ESPACIAL.....	70
4. METODOLOGIA.....	73
4.1. A CIDADE DE ARAÇARIGUAMA	76
4.1.1 COLETA DE DADOS	77
4.2 A CIDADE DE SÃO CARLOS	79
4.2.1 COLETA DE DADOS	79

4.2.1.1. DIMENSIONAMENTO DA AMOSTRA	82
5. RESULTADOS	88
5.1. ESTUDO DE CASO: ARAÇARIGUAMA	88
5.1.1. MODELO DE REGRESSÃO PARA A CIDADE DE ARAÇARIGUAMA	88
5.1.2. REDES NEURAIIS PARA A CIDADE DE ARAÇARIGUAMA	95
5.1.3. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELOS DOIS MÉTODOS PARA A CIDADE DE ARAÇARIGUAMA	97
5.2. ESTUDO DE CASO: SÃO CARLOS	104
5.2.1. MODELO DE REGRESSÃO PARA A CIDADE DE SÃO CARLOS	104
5.2.2. APLICAÇÃO DAS REDES NEURAIIS ARTIFICIAIS PARA SÃO CARLOS	112
5.2.3. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELOS DOIS MÉTODOS PARA A CIDADE DE SÃO CARLOS	115
5.3. AVALIAÇÃO DO DESEMPENHO DAS REDES NEURAIIS	122
6. CONCLUSÕES	132
7. BIBLIOGRAFIA	138

LISTA DE FIGURAS

Figura 2.1 - Queda nos preços de residências e terrenos obtida pelo estudo de ABELSON (1997).....	25
Figura 3.1 - Modelo não linear de um neurônio.....	35
Figura 3.2 - Comportamento de classificação de dois padrões de uma rede com uma camada intermediária	39
Figura 3.3 - Topologia de uma rede <i>multilayer perceptron</i> com uma camada intermediária	40
Figura 3.4 - Diagrama esquemático do treinamento de uma rede MLP	43
Figura 3.5 - Diagrama esquemático da fase de teste de uma rede MLP.....	44
Figura 3.6 – Regressões amostral e populacional.....	49
Figura 3.7 – Diagrama esquemático de um SIG	66
Figura 3.8 – Modelo matricial ou <i>raster</i>	67
Figura 3.9 – Modelo vetorial.....	68
Figura 4.1 - Distribuição espacial dos lotes pesquisados na cidade de Araçariguama..	78
Figura 4.2 - Distribuição espacial dos lotes pesquisados na cidade de São Carlos.....	87
Figura 5.1 -Erros relativos obtidos para os dois métodos e o primeiro conjunto de dados para a cidade de Araçariguama.....	98
Figura 5.2 - Erros relativos obtidos para os dois métodos e o segundo conjunto de dados para a cidade de Araçariguama.....	98
Figura 5.3 - Erros relativos obtidos para os dois métodos e o terceiro conjunto de dados para a cidade de Araçariguama.....	99
Figura 5.4 – Distribuição espacial dos erros para o modelo de regressão e o primeiro grupo de dados de Araçariguama.....	101
Figura 5.5 – Distribuição espacial dos erros para as RN e o primeiro grupo de dados de Araçariguama.	101
Figura 5.6 – Distribuição espacial dos erros para o modelo de regressão e o segundo grupo de dados de Araçariguama	102
Figura 5.7 – Distribuição espacial dos erros para as RN e o segundo grupo de dados de Araçariguama	102
Figura 5.8 – Distribuição espacial dos erros para o modelo de regressão e o terceiro grupo de dados de Araçariguama	103
Figura 5.9 – Distribuição espacial dos erros para as RN e o terceiro grupo de dados de Araçariguama	103
Figura 5.10 -Erros relativos obtidos para os dois métodos e o primeiro conjunto de dados para a cidade de São Carlos.....	116

Figura 5.11 - Erros relativos obtidos para os dois métodos e o segundo conjunto de dados para a cidade de São Carlos.....	116
Figura 5.12 - Erros relativos obtidos para os dois métodos e o terceiro conjunto de dados para a cidade de São Carlos.....	117
Figura 5.13 - Distribuição espacial dos erros para o modelo de regressão e o primeiro grupo de dados de São Carlos.....	119
Figura 5.14 - Distribuição espacial dos erros para as RN e o primeiro grupo de dados de São Carlos	119
Figura 5.15 - Distribuição espacial dos erros para o modelo de regressão e o segundo grupo de dados de São Carlos	120
Figura 5.16 - Distribuição espacial dos erros para as RN e o segundo grupo de dados de São Carlos	120
Figura 5.17 - Distribuição espacial dos erros para o modelo de regressão e o terceiro grupo de dados de São Carlos	121
Figura 5.18 - Distribuição espacial dos erros para as RN e o terceiro grupo de dados de São Carlos	121
Figura 5.19 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o primeiro conjunto de dados e a cidade de Araçariguama	127
Figura 5.20 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o segundo conjunto de dados e a cidade de Araçariguama.....	128
Figura 5.21 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o terceiro conjunto de dados e a cidade de Araçariguama.....	128
Figura 5.22 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o primeiro conjunto de dados e a cidade de São Carlos	128
Figura 5.23 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o segundo conjunto de dados e a cidade de São Carlos.....	129
Figura 5.24 - Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o terceiro conjunto de dados e a cidade de São Carlos.....	129
Figura 5.25 - Distribuição espacial dos erros obtidos para as RN e o primeiro grupo de dados de São Carlos quando o valor total do lote foi utilizado	130
Figura 5.26 - Distribuição espacial dos erros obtidos para as RN e o segundo grupo de dados de São Carlos quando o valor total do lote foi utilizado	131
Figura 5.27 - Distribuição espacial dos erros obtidos para as RN e o terceiro grupo de dados de São Carlos quando o valor total do lote foi utilizado	131

LISTA DE TABELAS

Tabela 3.1 - Tabela de Análise de Variância	53
Tabela 4.1 - Classificação dos bairros da cidade de São Carlos pela classe de preço A	83
Tabela 4.2 - Classificação dos bairros da cidade de São Carlos pela classe de preço B	83
Tabela 4.3 - Classificação dos bairros da cidade de São Carlos pela classe de preço C	84
Tabela 4.4 - Características das categorias	86
Tabela 5.1 - Melhores modelos de regressão ajustados, considerando o preço por m ² para a cidade de Araçariçuama.....	90
Tabela 5.2 - Valores de π_{ij} obtidos para as colunas da matriz X'X para Araçariçuama.....	91
Tabela 5.3 - Valores dos R^2_{ij} obtidos para cada par de variáveis para a cidade de Araçariçuama.....	91
Tabela 5.4 - Modelo de Regressão Ridge obtido para Araçariçuama utilizando o valor do m ² como variável dependente	92
Tabela 5.5 - Modelo de Regressão obtido para Araçariçuama utilizando o valor do m ² como variável dependente após a retirada da variável ln(testada)	93
Tabela 5.6 - Resultados das melhores simulações, usando todas as variáveis para a cidade de Araçariçuama.....	96
Tabela 5.7 - Relevância de cada variável para a cidade de Araçariçuama.....	96
Tabela 5.8 - Erros e R ² obtidos para os dois modelos e os três conjuntos de dados para Araçariçuama	97
Tabela 5.9 - Proporção de erros relativos nos <i>clusters</i> encontrados com os dois modelos para os três conjuntos de dados para Araçariçuama	100
Tabela 5.10 - Melhores modelos de regressão ajustados, considerando o preço/m ² para São Carlos	107
Tabela 5.11 - Valores de π_{ij} obtidos para as colunas da matriz X'X para São Carlos	108
Tabela 5.12 - Valores dos R^2_{ij} obtidos para cada par de variáveis para a cidade de São Carlos.....	109
Tabela 5.13 - Modelo de Regressão <i>Ridge</i> obtido para São Carlos utilizando o valor do m ² como variável dependente	110
Tabela 5.14 - Modelo de Regressão obtido para São Carlos utilizando o valor do m ² como variável dependente após a retirada das variáveis testada, calçada, edificação e água	110

Tabela 5.15 - Resultados das melhores simulações, usando todas as variáveis e o preço do m ² para São Carlos.....	112
Tabela 5.16 - Resultados das melhores simulações, usando todas as variáveis e o ln do preço do m ² para São Carlos.....	113
Tabela 5.17 - Relevância de cada variável para a cidade de São Carlos.....	114
Tabela 5.18 - Erros e R ² obtidos para os dois modelos e os três conjuntos de dados para São Carlos	115
Tabela 5.19 - Proporção de erros relativos nos <i>clusters</i> encontrados com os dois modelos para os três conjuntos de dados para São Carlos	118
Tabela 5.20 - Resultados obtidos pelas Redes Neurais quando o valor total do terreno foi utilizado como saída para a cidade de Araçariguama.....	123
Tabela 5.21 - Relevância de cada variável para a cidade de Araçariguama quando o preço total do terreno foi utilizado.....	124
Tabela 5.22 - Resultados obtidos pelas Redes Neurais quando o valor total do terreno foi utilizado como saída para a cidade de São Carlos.....	124
Tabela 5.23 - Relevância de cada variável para a cidade de São Carlos, quando o preço total do terreno foi considerando.....	126

RESUMO

Um dos problemas freqüentes em modelos de avaliação de imóveis é identificar quais de suas características devem ser levadas em consideração e o quanto cada uma destas influencia no valor final das propriedades. Além disso, é necessário um critério de avaliação bem estruturado, baseado em modelagem matemática adequada. A partir das constatações acima, foram estabelecidos os objetivos deste trabalho: após identificar as principais variáveis que interferem no valor das propriedades, avaliar o uso de Redes Neurais Artificiais para fins de avaliação e estudar a influência de uma medida de acessibilidade no valor de terrenos urbanos.

Quanto às variáveis a serem empregadas nos modelos de avaliação, chegou-se a conclusão que um banco de dados misto, onde tanto variáveis de natureza espacial quanto física pudessem ser incluídas, parecia ser uma opção interessante. Desta forma, após a inclusão no banco de dados de uma variável de natureza espacial, a distância ao centro da cidade, este trabalho comparou dois métodos de avaliação: as Redes Neurais Artificiais e o modelo de regressão múltipla, este último muito usado na prática. Foram abordados dois estudos de caso, as cidades de Araçariguama e São Carlos. A primeira é uma cidade dormitório, de pequeno porte (cerca de 6000 habitantes), localizada nas proximidades da capital do estado, São Paulo. A segunda, por sua vez, é uma cidade de porte médio (cerca de 160000 habitantes), localizada no centro do estado e pólo industrial e tecnológico. A escolha destas cidades ofereceu a oportunidade de estudar a influência de uma variável como a acessibilidade em contextos diferentes.

Os resultados obtidos para Araçariguama indicaram que a medida de acessibilidade empregada, distância ao centro, era uma das variáveis mais importantes na formação do preço de propriedades. Quanto aos resultados obtidos pelos modelos empregados, pôde-se observar que, ao utilizar regressão múltipla, o efeito da variável distância ao centro não pode ser estudado sozinho, pois esta variável interage com a área. A utilização de Redes Neurais, por sua vez, também forneceu estimativas adequadas de valor, verificando-se através dela que a acessibilidade apresentou um peso superior a 34% no valor final. Ao se comparar os dois métodos pode-se observar que as Redes Neurais (RN) demonstraram um desempenho superior, quando este foi avaliado pelo valor do erro relativo total. Com o objetivo de analisar a distribuição espacial dos erros, estes foram agrupados em cinco *clusters*, podendo-se constatar que os maiores erros fornecidos pelas RN se concentraram em um único bairro.

Na análise para São Carlos pôde-se verificar também, através dos resultados obtidos por ambos os métodos, que a distância ao centro foi um dos fatores preponderantes na avaliação dos imóveis. A análise da distribuição espacial dos erros apontou uma concentração de erros maiores em um dos bairros, o que pôde ser observado para os dois métodos empregados e dois dos três conjuntos de dados. Um fato que chamou a atenção foi o de que para Araçariguama, que é uma cidade de porte menor, a importância relativa da variável acessibilidade foi maior que para a outra cidade.

ABSTRACT

A common problem in the use of land valuation models is the identification of the real estate features that should be incorporated in the models and how they influence the final property price. In addition, a well structured approach based in consistent mathematical models is also required. Based on the aforementioned assertions, the following objectives have been drawn for this work: after identifying the main variables that have a strong influence on land values, the use of Artificial Neural Networks (ANN) for land valuation have be tested and the influence of an accessibility measure on urban land values have been studied.

Regarding the variables that should be part of the valuation models, we reached the conclusion that a mixed database containing physical and spatial attributes seemed to be an interesting option for this sort of problem. Therefore, after the addition of a spatial variable, the distance to the city center, to our database, two valuation methods have been compared: the ANN approach and a multiple regression model, the latter quite common in practice. Two case studies have been then analyzed: the cities of Araçariguama and São Carlos. The first one is a small bedroom town (around 6,000 inhabitants) not far from the state capital, the city of São Paulo. The second one is a medium-sized city (around 160,000 inhabitants) located in the middle of the state and a technological and industrial center. The particularities of these two cities made possible a comparison of the influence that such a variable as accessibility could have on the land values under two different conditions.

The results obtained for the city of Araçariguama indicated that the accessibility measure used, *the distance from the city center*, was one of the main variables influencing land prices. Although both models gave good estimates, their results were not exactly the same. While the influence of the variable *distance to the city center* could not be individually taken in the multiple regression model, because of its interaction with the variable *area*, the same variable has a strong weight on the ANN model, in which it appears as responsible for over 34% of the land value. The ANN performed better in a direct comparison of the two approaches, specially when looking to the total relative error. With the purpose of analyzing the spatial distribution of the estimation errors, they have been grouped into clusters, which have stressed that the worst cases are concentrated in a specific area of the city.

Both methods showed that the distance to the city center has a strong influence on land values also in the city of São Carlos. The highest estimation errors were also concentrated in a specific neighborhood for two out of three data sets in both valuation methods. Another interesting outcome is the fact that the relative weight of the accessibility variable used was higher in Araçariguama than in São Carlos, although the former city is smaller than the latter.

1. INTRODUÇÃO

Atualmente, a forma pela qual os impostos sobre a propriedade imóvel urbana devem ser cobrados é uma matéria de grande discussão (FERRAZ & SILVA, 1990; GONÇALVES, 1988; GONÇALVES, 1990; LONGO, s. d.; LONGO & LIMA, s.d.).

Em certa medida, o processo de avaliação de um imóvel para fins de tributação é semelhante ao processo de avaliação para fins de negociação (compra e venda). No primeiro caso, no entanto, em muitas cidades brasileiras o imposto é quantificado pela medida da área e localização da propriedade urbana, ignorando-se outros fatores tais como: acesso aos meios de transporte, condições da construção, infra-estrutura urbana, renda familiar etc. Com o objetivo de suprir esta lacuna, têm sido elaboradas propostas (FERRAZ, 1990; FERRAZ & SILVA, 1992; RAIA JR., 1995; SILVA, 1993; SILVA, 1998) de cobrança de impostos com o intuito de criar uma sistemática de tributação mais justa da propriedade urbana, onde a valorização de terras ociosas devido à acessibilidade aos meios de transporte e infra-estrutura levada a terras mais distantes deve ser levada em conta.

O tipo de variável a ser empregada nos modelos é um dos problemas em modelos de avaliação. Quais características devem ser levadas em consideração e o quanto cada uma destas influencia o valor final das propriedades são problemas freqüentes.

Desta forma, além de variáveis físicas, selecionadas de acordo com modelos encontrados na bibliografia revisada, uma variável de natureza espacial, a acessibilidade, também será incluída nos modelos de avaliação com o objetivo de

medir o impacto que esta pode causar nos preços de terrenos localizados em duas cidades com diferentes contextos regionais. Além de medir a influência da acessibilidade na valorização de imóveis urbanos, este trabalho pode servir como um auxiliar no processo de avaliação de propriedades urbanas, o que poderia ser de grande valia tanto para os órgãos públicos (no que diz respeito à cobrança de impostos) quanto para o setor privado.

Porém, a escolha adequada de variáveis não é o único problema na avaliação de imóveis, pois também é necessário um critério de avaliação bem estruturado, baseado em modelagem matemática adequada. Embora vários modelos de avaliação tenham sido propostos, a maioria das prefeituras brasileiras ainda baseia seus métodos de avaliação para cobrança de tributos em Plantas Genéricas de Valores (PGV), o que pode acarretar distorções e impostos cobrados injustamente quando sua construção e/ou atualização não são feitas de maneira adequada. Pelo sistema de cobrança vigente em muitas cidades, duas propriedades situadas no mesmo setor da planta, uma próxima do centro comercial do bairro, por exemplo, e a outra localizada mais distante, podem pagar o mesmo valor de imposto territorial. Entre estes casos, inclui-se a cidade de São Carlos, um dos objetos de estudo deste trabalho, cuja PGV não sofreu reformulação durante muitos anos, período em que a avaliação de terrenos com fins de tributação se dava apenas por atualizações monetárias.

Dentre os critérios de modelagem matemática utilizados para avaliação imobiliária, pode-se citar alguns métodos estatísticos. Embora eficientes, desde que utilizados por pessoal capacitado, em geral, eles podem estar comprometidos por vários fatores, como por exemplo: complexidade dos modelos, dificuldade de implementação, excesso de variáveis envolvidas e desconhecimento da relação entre estas variáveis. Incluem-se aqui, os modelos de Regressão Linear Múltipla, onde descontinuidades e não linearidades podem destruir a eficácia de toda uma análise.

Por outro lado, a escolha de Redes Neurais para o estudo destas estimativas, apresenta, dentre outras vantagens: a própria rede desenvolve as relações entre as variáveis, há tolerância no que diz respeito a erros ou distúrbios de

entrada (pela sua arquitetura, a própria rede é capaz de proporcionar uma saída correta para uma entrada deste tipo), além da não necessidade de determinadas condições impostas pelos métodos estatísticos.

Uma Rede Neural é baseada na estrutura biológica do cérebro humano através de uma analogia com os neurônios e tem como objetivo fornecer subsídios para que o computador "pense" sozinho e, a partir de um conjunto de situações conhecidas, tome determinadas decisões. Baseados nesses conhecimentos, cientistas procuraram desenvolver esquemas que possibilitassem a "construção" de neurônios artificiais, a fim de que estes pudessem efetuar diversas tarefas, tais como reconhecer imagens, aprender e memorizar associações. Para que o aprendizado ocorra é preciso que a rede seja treinada com situações conhecidas. A partir daí, ela "cruzará" estas informações que recebeu e será capaz de fornecer determinadas respostas.

Dentre as propriedades de um sistema neural, pode-se citar ainda a *convergência*, que diz respeito à capacidade da rede de aprender um determinado conjunto de dados definido no procedimento de treinamento e a *generalização*, que diz respeito à habilidade da rede de reconhecer certos padrões.

Atualmente, as Redes Neurais Artificiais (ANN) vêm sendo aplicadas nos mais diversos campos e em grande parte dos casos estudados têm apresentado resultados bastante favoráveis. Entre os campos de aplicação desta técnica, pode-se citar a medicina (SALCHENBERGER et al, 1997), a engenharia de transportes (DIA & ROSE, 1995; BIELLI & REVERBERI, 1996 e MUSSONI et al., 1996, RODGHER et al., 1997), a pesquisa operacional (KWON et al., 1995) e a educação (HARDGRAVE et al., 1994; FLITMAN, 1997), planejamento (WYATT, 1995, 1996 e 1997).

1.1. OBJETIVOS

Diante do que se expôs, o presente trabalho visa:

- Identificar as variáveis que interferem na formação de valor das propriedades;
- Avaliar o uso de Redes Neurais Artificiais para fins de avaliação e
- Estudar a influência de uma medida de acessibilidade no valor de terrenos urbanos.

1.2. MÉTODO

Para atingir os objetivos deste trabalho, os seguintes procedimentos serão adotados:

- i. Revisão bibliográfica relativa à atual situação da avaliação imobiliária nacional e internacional, no que tange a variáveis mais empregadas e métodos;
- ii. Revisão bibliográfica sobre Redes Neurais, Regressão e Sistemas de Informação Geográfica, ferramentas que serão utilizadas para este estudo;
- iii. Com base nas variáveis mais utilizadas na literatura (item i.), executar a coleta de dados;
- iv. Estimar os valores imobiliários e os erros obtidos, utilizando Regressão Múltipla e Redes Neurais;
- v. Verificar o peso de cada variável, em especial da acessibilidade, no valor estimado final para a Regressão;
- vi. Verificar a possibilidade de utilizar Redes Neurais para avaliar o impacto de cada variável, em especial da acessibilidade, no valor estimado final;
- vii. Comparar os resultados fornecidos por ambos os métodos, numérica e espacialmente.

1.3. ESTRUTURA DO TRABALHO

Para alcançar os objetivos almejados, este trabalho se divide em cinco capítulos, além desta introdução, o primeiro dos quais se inicia com um levantamento bibliográfico. Nele, a atual situação da avaliação de imóveis no Brasil é comentada. Na seqüência, alguns aspectos da avaliação coletiva de imóveis serão apresentados, assim como alguns modelos de avaliação nacionais e internacionais, acompanhados das respectivas variáveis envolvidas. O capítulo finaliza comentando aspectos relativos à cobrança de impostos sobre a propriedade imobiliária sob o ponto de vista de diferentes autores.

No terceiro capítulo será feita uma introdução à evolução da Inteligência Artificial, relativa, principalmente, à área de Redes Neurais. A ênfase será no modelo empregado neste trabalho, ou seja, a Rede Multilayer Perceptron. O capítulo continua apresentando um resumo sobre Regressão Múltipla, que foi a técnica utilizada para a comparação dos resultados. Finalizando este capítulo, alguns comentários sobre os Sistemas de Informação Geográfica, cuja ferramentas foram utilizadas neste trabalho para tratar cada propriedade de forma diferenciada, serão traçados.

O quarto capítulo refere-se à metodologia aplicada. Neste item, serão explicados os procedimentos adotados no decorrer desta pesquisa. Haverá também uma descrição dos objetos de estudo e da forma de coleta de dados empregada.

Na seqüência, o quinto capítulo consta de dois estudos de caso, onde os resultados obtidos do emprego das duas técnicas serão apresentados. Finalmente, o sexto capítulo apresenta algumas conclusões, seguido pelas referências bibliográficas utilizadas na pesquisa.

2. AVALIAÇÃO DE IMÓVEIS

Este capítulo inicia com uma consideração sobre a acessibilidade e o valor do solo, do ponto de vista da teoria econômica. A segunda parte é destinada a um breve levantamento da situação da avaliação de imóveis atualmente em vigor no Brasil. Numa próxima etapa, alguns aspectos da avaliação coletiva de imóveis serão discutidos. As duas seções posteriores mostram alguns modelos utilizados por autores nacionais e internacionais, que incluem tanto variáveis de caráter físico quanto espacial. Finalizando o capítulo, aspectos da taxaçoão sobre o valor das propriedades sob o ponto de vista de diferentes autores serão brevemente discutidos.

2.1. ALGUMAS CONSIDERAÇÕES SOBRE ACESSIBILIDADE E O VALOR DO SOLO

Antes de comentar a relação entre a acessibilidade e o preço do solo, algumas considerações teóricas com respeito à evolução histórica do valor da terra devem ser feitas. RAIJA JR. (1997), com base no texto de NIGRIELLO (1977)¹, aborda a história da teoria do valor, classificando-a segundo o ponto de vista das escolas clássica, marginalista e marxista.

A *teoria clássica* considera que o valor do solo é resultante, de forma indireta, do rendimento que o solo pode propiciar por meio de seu cultivo. Este

¹ NIGRIELLO, A. 1977. *O valor do solo e sua relação com a acessibilidade*. São Paulo. 238p. Dissertação (mestrado) – COPPE, Universidade Federal do Rio de Janeiro.

rendimento é dado, de maneira indireta, pela diferença entre o montante percebido com a venda dos produtos agrícolas e o investimento realizado na produção dos mesmos. Devido ao quadro social e econômico da época em questão, os pensadores desta teoria deram maior ênfase à renda do solo rural pois a renda do solo urbano era, até então, pouco significativa. Porém, nesta fase, já se admitia que a localização da propriedade tinha influência sobre sua renda.

A fase de transição entre as escolas clássica e marginalista iniciou-se com pensadores que defendiam as idéias de que o solo empregado em várias funções urbanas também gerava rendas.

Assim, segundo a *teoria marginalista*, o valor de um bem é baseado na utilidade que uma quantidade marginal deste bem possa propiciar, que não coincide, necessariamente, com seu custo de produção. Por esta teoria, o solo é o próprio bem a ser comercializado e seu valor depende do total da renda futura que este pode propiciar.

Numa posição intermediária entre estas duas escolas, encontram-se as idéias de pensadores como Von Thunen e Hurd. O primeiro apregoava que a renda do solo agrícola era a diferença entre o preço final do produto no mercado e a somatória do seu custo de produção e de transporte. O segundo, desenvolveu um modelo onde se concluía que à medida em que a área urbana se amplia, o nível de renda do solo localizado nas proximidades da zona central cresce, assim como as diferenças entre o centro e a periferia.

Uma relação entre o valor do solo e os custos de transporte vem das idéias de Haig, para o qual o *custo de fricção*, representado pelo custo de aluguel mais o custo de transporte, não é constante e varia de lugar para lugar, permitindo afirmar que os aluguéis de imóveis próximos à zona central são mais baixos porque os custos de transporte também o são.

Do ponto de vista de Ratcliff, três fatores influenciam o valor do solo, a saber:

- Físicos: a imobilidade do solo não permite o deslocamento do bem de um lugar para outro para atender à demanda e sua indestrutibilidade sempre garante a ele um valor;

- **Localização:** são aqueles onde a localização de diversos usos do solo é produto da concorrência do mercado – melhor emprego do solo, maior valor. Exemplos destes fatores são o sistema de transporte, cuja melhoria pode levar a um aumento do valor em regiões onde há melhor acessibilidade, o prestígio e as características sociais.
- **Regulamentação:** considera as restrições legais, as normas e leis de zoneamento etc.

Wendt, embora reconhecendo que os valores dos imóveis são incrementados com a melhoria do sistema de transporte, contesta a importância da acessibilidade ao centro como fator principal de localização das funções urbanas e alega que esta deve ser considerada com relação à toda região urbana e não só com relação à área central.

A terceira teoria, a *marxista*, afirma que os principais fatores vinculados aos preços de mercado de imóveis urbanos, em determinado instante, são a acessibilidade aos centros urbanos, o nível de renda da demanda por determinado tipo de imóvel, o coeficiente de ocupação do solo, o entorno físico e os serviços e equipamentos locais. Assim, de acordo com estes mecanismos, se configuram as alterações no meio urbano, inclusive as melhorias no sistema de transporte.

Vieille afirma que o valor do solo é resultante da atividade coletiva, ou seja, do agrupamento dinâmico dos indivíduos e classifica a mais valia do solo em dois componentes: a mais valia comercial e a mais valia geral. A primeira se concentra no centro principal, nos secundários e onde é comum a reunião de população e o intercâmbio, afetando o valor de terrenos juntos às avenidas e ruas que cruzam estes centros e é desfrutada pelos proprietários de terrenos e comerciantes. A segunda refere-se à área mais periférica e nela os valores de terrenos representam a capitalização das rendas relativas de localização, que seus proprietários requerem em função das condições de acesso ao centro. Desta forma, os terrenos próximos ao centro são os mais disputados em virtude de sua localização e conseqüentemente alcançam os valores mais elevados.

Como se observa nas teorias clássicas comentadas acima, apesar da contribuição de outros fatores econômicos e sociais, não se pode negar a importância da acessibilidade no valor do solo. Este fato pode ser verificado em alguns estudos sobre avaliação realizados não só no Brasil mas também em outros países e que serão discutidos mais adiante.

2.2. SITUAÇÃO ATUAL DA AVALIAÇÃO DE IMÓVEIS

Uma forma precisa de estimar o valor da propriedade urbana tem sido almejada não só pelo setor público, com o intuito de quantificar devidamente o valor de seus impostos (IPTU – Imposto Predial e Territorial Urbano e ITBI – Imposto Sobre Transações de Bens Imobiliários), mas também pelo setor privado, que não quer nem pagar e nem cobrar mais pelas terras que adquire ou vende.

Numa tentativa de estabelecer certas regras e critérios para avaliação, a Associação Brasileira de Normas Técnicas (ABNT - 1989), propõe uma “Norma Brasileira para Avaliação de Imóveis Urbanos”, a NBR 5676. Tal norma estabelece três níveis distintos de rigor para a realização de uma avaliação, com graus crescentes de refinamento matemático e fundamentação, resumidos a seguir:

- **expedito:** baseado na subjetividade, não utiliza qualquer instrumento matemático para a estimação de valor.
- **de precisão:** utilização de, no mínimo, cinco dados da propriedade para a formação do valor. O tratamento dispensado aos elementos da amostra “... deve ser feito através de estatística descritiva, homogeneizando-se os elementos observados, quando não perfeitamente comparáveis” e “... são admitidas as homogeneizações feitas através de fatores ou ponderações empíricas, porém, consagradas, que consideram as diferenças entre os elementos de seus atributos”.
- **de precisão rigorosa:** a avaliação deve ser feita “... através de metodologia adequada, com grande isenção de subjetividade...” e “... o tratamento para alcançar a convicção de valor deve ser baseado em

processos de inferência estatística que permitam calcular estimativas não tendenciosas de valor, estabelecendo intervalos de confiança para elas, além de submetê-las a testes de hipóteses...”.

A maioria dos trabalhos técnicos publicados sobre avaliação têm sugerido a utilização da avaliação de precisão rigorosa (FIKER, 1993 e 1989; MEDEIROS JR. & FIKER, 1996; MOREIRA, 1997). Em grande parte destes, a Análise de Regressão Múltipla é utilizada como ferramenta para a estimação de valor. Porém, embora eficientes, os modelos deste tipo esbarram em algumas limitações, como por exemplo, existência de descontinuidades, desconhecimento de relações entre as variáveis envolvidas e obediência de suposições quanto à distribuição dos erros. Por outro lado, está crescendo o uso de modelos de Inteligência Artificial, em particular, Redes Neurais, em substituição a modelos estatísticos. Dentre as áreas que estão utilizando ferramentas deste tipo, pode-se citar, por exemplo, psicologia, fonoaudiologia, pedagogia, medicina e engenharia, inclusive na área de transportes. Dada sua capacidade de reconhecimento e generalização, estes modelos reagem bem e fornecem estimativas adequadas na existência de ruídos ou quando falta algum tipo de dado, classificando aquele conjunto de entrada na classe em que ele melhor se enquadre. Como outra vantagem, a rede é capaz de aprender e, assim, fornecer saídas satisfatórias para novas entradas.

Além do nível de rigor, a Norma NBR 5676 também indica variáveis formadoras de valor para a avaliação de terrenos, que são: caracterização física (relevo, solo, subsolo e ocupação); acessos, serviços e melhoramentos públicos; utilização atual e potencial, legal e econômica; classificação do imóvel. Para terrenos loteados, também devem ser consideradas as seguintes influências: 1) área e profundidade e 2) frente.

Atualmente, na maioria das prefeituras brasileiras os sistemas de avaliação são baseados em plantas de valores, que nem sempre são elaboradas de acordo com os métodos recomendados por especialistas no assunto. Nos casos mais extremos, onde as plantas não foram elaboradas ou mesmo atualizadas com o

devido rigor, pode acontecer de lotes situados no mesmo setor apresentarem o mesmo valor por metro quadrado, obtido geralmente através de avaliação expedita, independentemente de suas características individuais. Esta “generalização” pode provocar grandes distorções de valor. Somente para exemplificar, algumas observações, dentre muitas, podem ser citadas com relação a situações deste tipo:

- Lotes situados a menos de 100 m um do outro e com as mesmas características físicas e espaciais, podem apresentar valores bastante diferentes por estarem localizados em setores vizinhos.
- Dois lotes situados no mesmo bairro, um localizado numa região plana e outro em um lugar cuja topografia seja irregular e, portanto, apresentando menor valor de mercado, serão avaliados pelo mesmo valor por metro quadrado.
- O valor do metro quadrado de um bairro que recebeu um centro de recreações aumentou. Conseqüentemente, propriedades situadas no mesmo bairro, mas distantes de tal centro, sofrerão uma “valorização”. Porém, propriedades localizadas num setor vizinho e que estejam a uma distância menor do mesmo centro de recreações, terão seus valores inalterados, embora sofram mais a influência deste.

Como se não bastassem os fatores já mencionados, a reavaliação cadastral não é realizada anualmente. Em lugar desta, são usados índices genéricos de correção monetária, que não seguem o mercado imobiliário. Assim, imóveis localizados em áreas que sofreram uma grande valorização continuam pagando menos pelos tributos incidentes sobre a propriedade. Um fator agravante para esta situação é quando a valorização se dá por conta de melhorias financiadas pelo próprio setor público (pavimentação, por exemplo).

Apesar de todos os problemas mencionados, poucas pessoas têm se dedicado a estudar a valorização de cada propriedade de forma diferenciada. Além disso, em grande parte dos estudos propostos, a influência da acessibilidade tem

sido relegada a segundo plano, embora muitos autores tenham comprovado sua importância quando se considera a valorização de imóveis. Os poucos autores que tem considerado alguma medida deste tipo, usam, em geral, a distância até o centro, o que, para cidades de porte pequeno pode ser suficiente. Por outro lado, cidades grandes possuem bairros completamente auto-suficientes em termos de comércio, escolas, hospitais e até mesmo de lazer, fazendo com que a principal preocupação dos compradores que desejem morar em determinado lugar seja a facilidade de acesso ao local de trabalho, caso este se situe em uma região distante. Nestes casos, talvez seja interessante estudar também a acessibilidade aos meios de transporte (por exemplo, linhas de transporte coletivo) como fator de valorização.

Finalmente, vale observar que, neste trabalho, optou-se por trabalhar apenas com a avaliação de terrenos, pois, além do menor número de variáveis envolvidas, a coleta de dados é mais rápida e fácil, principalmente quando feita *in loco*. Além disso, o que diferencia o valor de propriedades localizadas em bairros diferentes é o preço do terreno, uma vez que os gastos com a construção só dependem do padrão e são os mesmos, independentemente do lugar onde se deseja edificar. GONZÁLEZ & FORMOSO (1994) concluíram em um trabalho realizado para a cidade de Porto Alegre, que “o preço dos imóveis modifica-se substancialmente de acordo com sua localização, enquanto que o custo de construção é praticamente o mesmo em todas as regiões da cidade”. Portanto, as características da região onde a propriedade se situa são uma influência bastante significativa no valor destas.

2.3. AVALIAÇÃO COLETIVA DE IMÓVEIS

As avaliações coletivas ou em massa buscam a determinação de valores que servem como base de cálculo para fixação de impostos ou contribuições. Antes da aplicação da metodologia para avaliação coletiva de imóveis, o código tributário do município deve ser estudado com atenção. Além disso, as regras do código tributário municipal têm que estar em conformidade com a Constituição Federal. Este é o primeiro obstáculo a ser enfrentado na implantação

de uma nova Planta Genérica de Valores (PGV), pois o imposto só pode ser majorado através de lei, ressalvado o princípio da anterioridade (qualquer mudança na base de cálculo ou valor do tributo deve ser apresentada até o final do ano anterior), salvo quando o valor for corrigido monetariamente através de decreto (GONZÁLEZ, 1996a). A utilização de correção monetária tem sido aceita pacificamente pela Justiça. Porém, o aumento do imposto acima da inflação do período, o que pode ocorrer em alguns casos depois da aplicação de uma nova PGV, tem sido considerado como uma majoração de tributos. Diante disso, antes de mais nada, a Planta Genérica de Valores tem que ser elaborada no ano anterior ao que se deseja aplicá-la.

Dentre as vantagens da elaboração, atualização e divulgação periódica de uma Planta Genérica de Valores (PGV), por parte do Poder Público Municipal, pode-se citar as seguintes (MÖLLER, 1995):

- A planta serve de base para cálculo dos impostos, que incidem sobre a propriedade urbana e que são de responsabilidade das prefeituras Municipais (IPTU e ITBI);
- A redistribuição da carga fiscal é permitida através da redução da diferença entre o valor fiscal e o de mercado, promovendo, desta forma, justiça social;
- Auxílio no planejamento urbano, permitindo a previsão de custos de desapropriação nas obras públicas, bem como o retorno provável de tais investimentos (contribuição de melhoria);
- Retrata a realidade do mercado local à época da sua elaboração, fornecendo justas bases para as transações, reduzindo os riscos e com eles as margens de manipulação dos consumidores pelos agentes imobiliários.

Além disso, a utilização de Planta de Valores Genéricos de Terrenos permite comparar a evolução de valores nas zonas beneficiadas pelos melhoramentos públicos, bem como numa faixa de segurança ao redor da referida

zona, durante um período de observação. A análise das avaliações em situações distintas (antes e depois) das melhorias possibilita a verificação das variações de valor ou eventual estagnação de mercado.

Apesar de todas as vantagens apresentadas, colocar em prática uma nova PGV não é tão simples. Como já mencionado, o primeiro obstáculo é a Justiça, que não tem permitido reajustes acima da inflação do período. Porém, a correção monetária plena, aplicada em imóvel que sofreu desvalorização ou que não se valorizou na mesma proporção da inflação equívale, de certa forma, a uma majoração de tributos.

2.3.1. COLETA DE DADOS

Para a boa elaboração de uma Planta Genérica de Valores, a coleta de dados é um ponto importante a ser observado. Dentre as fontes de informação disponíveis, o ITBI (Imposto Sobre Transação de Bens Imobiliários) é uma das mais acessíveis para as prefeituras. Segundo MÖLLER (1995), se a escolha da fonte de dados for o valor deste imposto, alguns tipos de transação devem ser rejeitadas: transações entre pessoas jurídicas ou sócios, doações especialmente entre parentes e hipotecas ou empréstimos com garantia de bens imóveis e desapropriações.

Assim, em boa parte dos casos, considerar como verdadeiro valor da propriedade aquele registrado em cartório pode não ser uma decisão acertada, uma vez que as pessoas que negociam propriedades podem, de comum acordo, registrá-las abaixo do preço da negociação, justamente para fugir do imposto. Além disso, nem sempre há registro de escrituras e, às vezes, há um longo período de tempo entre a venda e o registro, fazendo com que o valor registrado esteja defasado em termos financeiros.

Por outro lado, os elementos obtidos através da análise de ofertas fornecem boas bases para aferição dos preços praticados pelo mercado. As ofertas podem ser coletadas de várias formas, a saber: em jornais, na forma de anúncios

classificados; em imobiliárias e/ou com corretores autônomos ou através de placas afixadas nos imóveis.

Quanto às variáveis, MÖLLER (1995) sugere que para a construção adequada de uma Planta Genérica, sejam levantadas as seguintes características dos terrenos: localização, dimensões e forma, aspectos físicos (topografia, pedologia, várias frentes, recuos), equipamentos urbanos (água, luz, esgotos, pavimentação, transporte coletivo, etc.), data do negócio e fonte de informação (número do registro, origem, informante, preços e condições de pagamento).

Quanto ao dimensionamento da amostra, ZANCAN (1996) sugere que este tamanho seja determinado em função de estudos similares. Em seu livro, ela cita alguns tamanhos amostrais que foram utilizados para avaliação de casas e apartamentos, segundo trabalho de GONZÁLEZ (1993), que utilizou 500 dados de apartamentos para aluguel na sua dissertação de mestrado e relata o número de imóveis que constituem a amostra em trabalhos semelhantes:

- *Kain & Quieley – do ano 1970 – 605 imóveis residenciais / aluguel*
- *Witte et al. – do ano 1974 – 500 imóveis/ aluguel*
- *Berry & Bednarz – do ano 1975 – 275 imóveis residenciais / venda*
- *Sunca – do ano 1977 – 270 imóveis*
- *Franchi – do ano 1991 – 85 imóveis residenciais / venda*
- *Dantas – do ano 1987 – 50 dados sobre terrenos em Recife na dissertação Avaliação de Glebas Inseridas na Malha Urbana*

Idealmente, o tamanho da amostra deve estar relacionado com as dimensões da cidade a ser utilizada no trabalho.

2.3.2. METODOLOGIA APLICADA NA CONSTRUÇÃO DE UMA PGV

Para entender a construção de uma PGV, algumas definições devem ser feitas a priori:

- **Zonas homogêneas:** são as zonas urbanas que apresentam a mesma evolução durante um determinado período de tempo, manifestada no porte e padrão construtivo das edificações, bem como pelo poder aquisitivo de seus habitantes. Existe uma tendência (não obrigatória) das zonas homogêneas acompanharem o zoneamento de uso e ocupação do solo;
- **Curvas isótimas:** lugar geométrico dos pontos de igual valor, curvas de níveis, permitindo a visualização de relevo de valores;
- **Pólos de valorização:** pontos ou linhas, a partir dos quais a distribuição de valor dentro do perímetro urbano de uma cidade se faz de forma decrescente. Quanto à importância de sua influência nas áreas vizinhas, classificam-se em pólos principais ou secundários. Normalmente os centros comerciais de uma cidade constituem-se em pólos principais pontuais. Nas cidades balneárias, em geral, a orla marítima é um pólo principal linear.

Quanto ao tipo de modelagem empregada, tem-se duas situações: os modelos predeterminados e os modelos inferenciais.

2.3.2.1. MODELOS PREDETERMINADOS

Por este método, a homogeneização se processa através de fatores de ponderação previamente determinados que corrigem as distorções entre terrenos no que tange à profundidade, topografia, pedologia, várias frentes (esquina ou encravamento), entre outros. O mesmo procedimento é feito em relação à existência, ou não, de equipamentos urbanos (água, energia elétrica, pavimentação, transporte coletivo, esgoto, entre outros), através de percentuais de valorização agregados ao terreno em cada tipo de serviço público.

O roteiro para elaboração da Planta Genérica de Valores de terrenos após a homogeneização da pesquisa de mercado, tem as seguintes etapas:

1. **Plotagem dos valores unitários e pólos de atração:** os valores unitários são plotados preliminarmente na face de quadra correspondente à sua localização. Na mesma planta são inseridos os pólos de valorização principais e secundários (pontuais e lineares). Os pólos comerciais constituem-se, em geral, nos principais pólos de valorização de uma cidade;
2. **Curvas e áreas isótimas:** pela análise gráfica das curvas isótimas, pode-se verificar a formação de áreas isótimas lineares ou setoriais, que são zonas delimitadas por linhas do mesmo valor, como acontece em alguns bairros residenciais;
3. **Complementação da pesquisa:** áreas não atingidas e dados discrepantes devem ser objetos de novas investigações de mercado;
4. **Discussão de valores:** valores obtidos devem ser debatidos com entidades representativas dos vários segmentos da comunidade imobiliária;
5. **Valores unitários finais:** após as etapas preparatórias, chega-se à Planta Genérica de Valores de terrenos, da qual são transcritos os valores de cada testada, quadra por quadra, através de códigos que identifiquem a zona ou setor, quadra e logradouro. Estes valores alimentam a memória do computador, quando for o caso, que os utiliza nos cálculos de valor dos terrenos.

Este tipo de modelo de construção de PGV's tem contra si o uso de parâmetros sem a estimação das suas relevâncias relativas na formação de valor. Ou seja, todos os fatores têm a mesma importância nesta composição de valor, quando se sabe que alguns são mais influentes que outros.

2.3.2.2. MODELOS PROBABILÍSTICOS OU INFERENCIAIS

Nos modelos inferenciais, a determinação do valor unitário básico de terrenos é feita através do Método Comparativo de Dados de Mercado, com base em Regressões Múltiplas. O método comparativo de dados de mercado é, segundo a NBR 5676:

“Aquele que define o valor pela comparação com dados de mercado assemelhados quanto às características intrínsecas e extrínsecas. As características e os atributos dos dados pesquisados que exercem influência na formação dos preços e conseqüentemente (sic), no valor, devem ser ponderados por homogeneização ou por inferência estatística, respeitados os níveis de rigor definidos neste norma. É condição fundamental para aplicação desse método a existência de um conjunto de dados que possa ser tomado, estatisticamente, como amostra do mercado imobiliário”.

Sendo assim, com base nas evidências oferecidas pelo mercado, detectadas pela pesquisa, os valores unitários de terreno em cada face de quadra são inferidos por um modelo estatístico.

Na elaboração de PGV, a variável dependente ou explicada é sempre o valor básico unitário do metro quadrado da face de quadra. As variáveis independentes podem ser qualitativas ou quantitativas.

Entre as quantitativas, destacam-se:

- Área, testada, profundidade;
- Distância às vias principais ou pólos de valorização.

Uma outra alternativa, que também pode ser adicionada como variável explicativa é o índice de aproveitamento do terreno. Um exemplo deste tipo de índice é a cota-terreno (CT), definida como o quociente do valor do terreno pela área construtível ideal.

Como variáveis qualitativas mais usuais, destacam-se:

- Topografia, pedologia;
- Várias frentes, esquina;
- Vizinhança, utilização;
- Localização, zoneamento (setor);
- Equipamentos públicos, urbanização;
- Data e elasticidade (oferta/transação).

A variável data da oferta ou da transação permite calcular a valorização imobiliária efetiva de terrenos setorialmente ou de toda a malha urbana, constituindo-se em um importante indicador para decisões técnicas ou políticas de investimento no município.

2.4. ALGUNS MODELOS DE AVALIAÇÃO NACIONAIS

Diante de um quadro de critérios de avaliação muito pobres que não é exclusivamente brasileiro, muitos pesquisadores têm se empenhado em desenvolver melhores métodos de estimação do valor de imóveis. Eles visam obter estimativas cujos valores se aproximem o máximo possível do valor real destas propriedades, sendo que cada modelo proposto utiliza os mais diversos tipos de variáveis. A seguir são resumidos alguns destes modelos.

BORGES (1975) fez um estudo a respeito da valorização de terrenos e levantou alguns fatores condicionantes para o diferencial de valor dos terrenos urbanos. Entre os principais fatores determinantes deste diferencial, ele destaca os

seguintes: distribuição espacial dos centros de emprego, a organização e tecnologia dos sistemas de transporte, as características sócio-econômicas da vizinhança, a complementaridade de uso do solo, a densidade demográfica, as características topográficas dos terrenos, as características dos lotes e a qualidade de infra-estrutura urbana. Ele fez uma análise econômica bastante completa da valorização de propriedades e propôs um modelo que usa acessibilidade (no caso, distância ao principal centro de empregos) para estudar a influência dos custos de transportes nesta avaliação.

GONZÁLEZ & FORMOSO (1994) propõem o uso do ITBI (Imposto Sobre Transações de Bens Imobiliários) para a estimação de valores de imóveis em Porto Alegre. Neste estudo, eles utilizam as seguintes variáveis independentes: ano de transação, mês de referência, área construída, idade, material de construção (madeira, alvenaria etc.), área do terreno, testada (frente), posição do terreno na quadra, renda da população no bairro, distância, presença de financiamento e ITBI. Para o modelo de estimação do valor de casas, em função destas variáveis, foi obtido um coeficiente de determinação igual a 0,73 e para os terrenos, um valor igual a 0,55.

Em GONZÁLEZ & FORMOSO (1995) e GONZÁLEZ (1995), além do ITBI, são consideradas como variáveis independentes para a avaliação de apartamentos: área, bairro, existência de garagem, distância ao centro, idade, “luxo” e mês de referência. O segundo trabalho, além de todas as variáveis mencionadas, considera as coordenadas (longitude e latitude) do imóvel. Além destes, outros autores trataram da avaliação de apartamentos, dentre os quais pode-se citar: HOCHHEIM et al. (1996), PERUZZO et al. (1996), GONZÁLEZ (1996b), PAULINO et al. (1995), BRANDÃO et al. (1995), BALARINE (1995) e ZANCAN & HEINECK (1994).

LAPOLLI et al. (1994) propuseram modelos para estimar o valor das propriedades para a cidade de Porto Alegre. Para tal trabalho, eles consideram uma série de variáveis para estimar os valores de casas, apartamentos, terrenos, salas comerciais, lojas e boxes de estacionamento. O modelo para avaliação de terrenos considera as seguintes variáveis: água, pavimentação, topografia, pólo de comércio

e de serviços, zona residencial, livre (sem construção), testada, área, data de comercialização (ou oferta) e fonte de informação. O valor do coeficiente de determinação obtido pelo modelo de regressão entre o preço de terrenos e as variáveis independentes mencionadas foi de 0,88.

VERTELO (1996) fez um estudo para avaliar bens imóveis da cidade de Governador Valadares. Ele desenvolveu modelos que servem para avaliar vários tipos de imóveis: casas, barracões, condomínios horizontais e verticais, terrenos, chácaras, galpões e garagens de edifícios. Para terrenos, as variáveis independentes consideradas foram: infra-estrutura urbana, topografia, área, testada e setor.

YUAÇA (1996) fez a proposta de utilização de um Estimador de Valor do Lote Urbano que toma como base o valor da Planta de Valores (PV) de Goiânia, com alguns ajustes. Inicialmente, foi feita uma suavização da transição entre uma PV e, logo após, vários índices foram determinados: SLOPE (verifica o grau de inclinação do lote), VIZINHOS ADJACENTES (identifica o padrão das casas dos vizinhos de muro), VIZINHOS PRÓXIMOS (identifica o padrão das casas das redondezas do lote em questão), OCORRÊNCIAS POSITIVAS E NEGATIVAS (determina a influência de cada ocorrência, como por exemplo, *shoppings*, escolas, esgotos abertos etc., sobre o lote) e LOGRADOUROS ESPECIAIS (verifica se o logradouro em questão é tratado de forma especial em termos de valor). Como resultado, alguns lotes apresentaram valores diferenciados dentro da mesma PV.

RAIA JR. et al. (1996a) propuseram um modelo que estima o preço de terrenos para a cidade de Araraquara através da distância ao centro. Porém, uma regressão simples entre os valores dos lotes e esta medida de acessibilidade forneceu um coeficiente de determinação muito pobre (0,37), indicando a necessidade de inclusão de outras variáveis no modelo.

Com o intuito de melhorar o modelo previamente citado, RAIA JR & SILVA (1996), propuseram um novo critério para avaliar o valor dos imóveis de Araraquara, incluindo novas medidas de acessibilidade. As medidas de acessibilidade usadas em tal estudo foram: distância entre o bairro onde se localiza o imóvel e o centro da cidade, distância do imóvel até o transporte coletivo, relação

entre as linhas de ônibus que servem o bairro e o seu sistema viário e pavimentação do sistema viário. Uma regressão feita entre os preços dos imóveis e estas medidas apresentou resultados piores que os anteriores (coeficiente de determinação igual a 0,26).

RAIA JR. et al. (1996b) incluíram o isolamento (proposto por DAVIDSON, 1995) como uma nova medida de acessibilidade ao seu modelo. Como consequência, a regressão entre todas as medidas de acessibilidade citadas anteriormente mais o isolamento produziu uma boa melhoria dos resultados, chegando à obtenção de um coeficiente de determinação igual a 0,48.

O modelo proposto por SILVA (1998) inclui outras duas medidas de acessibilidade para estimação do valor das propriedades em Araraquara: uma medida de acessibilidade do tipo gravitacional e a distância média entre zonas. A regressão feita entre os valores dos imóveis e todas as medidas de acessibilidade citadas forneceu uma melhoria do coeficiente de correlação (que pulou de 0,48 para 0,53). Este resultado mostrou que a acessibilidade é importante para a estimação dos preços destes imóveis, porém, não suficiente.

Um outro modelo proposto por BRONDINO & SILVA (1997) faz uso de dados armazenados em um Sistema de Informações Geográficas (SIG) para avaliar propriedades na cidade de Araçariçuama. A partir de certas características conhecidas (área, benfeitorias, localização dentro da quadra, forma, testada principal e distância ao centro), o valor do metro quadrado dos terrenos é obtido através de acréscimos ou decréscimos percentuais sobre o valor do metro quadrado de um lote padrão. O lote padrão é definido como aquele localizado a 500 m ou menos do centro da cidade, situado no meio da quadra, com muro e calçada, de forma regular e testada inferior a 20 m. Como resultado, grandes propriedades que estavam sendo subavaliadas tiveram seus valores corrigidos.

TOLEDO & DOMINGUES (1995) vão além da área dos municípios e propõem um modelo que avalia glebas e lotes em todo o estado de São Paulo. Para a formulação do modelo, foram levantadas 40 variáveis, sendo que as seguintes foram consideradas estatisticamente significantes, através dos testes pertinentes: população, distância ao centro do estado, distância ao pólo de atração, tendência de

uso de terreno - comercial, condição (gleba ou lote), região administrativa e mês de referência. O coeficiente de determinação obtido para a regressão entre o preço e as variáveis independentes foi de 0,72.

Como pôde ser notado pelas referências citadas anteriormente, os critérios de escolha de variáveis são os mais variados possíveis, de acordo com características da região de localização e do imóvel. Na grande maioria deles há o emprego de Análise de Regressão Múltipla, o que já era esperado, uma vez que este é o critério de avaliação mais usual e recomendado pela NBR 5676. Como exceções a esta regra, pode-se citar os modelos de BRONDINO & SILVA (1997) e GUEDES (1995). Neste último faz-se uso de recursos de Inteligência Artificial para estimar o valor de prédios comerciais no Rio de Janeiro. Neste modelo, o valor das propriedades foi estimado a partir das seguintes variáveis: nota atribuída ao padrão construtivo, idade aparente do prédio, existência de garagem, vizinhança do bairro de Botafogo ou não, localização no centro ou não e valor original. Como resultado, ele concluiu que os recursos de Inteligência Artificial se mostraram como uma ferramenta bastante poderosa para avaliações.

2.5. EXEMPLOS DE MODELOS DE AVALIAÇÃO DE IMÓVEIS ENCONTRADOS NA LITERATURA INTERNACIONAL

A maior parte dos trabalhos mencionados nesta seção fez uso de Análise de Regressão Múltipla. Pode-se verificar também que, em alguns países, a existência de bancos de dados antigos pôde promover um estudo detalhado da evolução dos preços ao longo do tempo. Outro fator que deve ser mencionado é que muitos dos autores aqui citados fazem uso de modelos hedônicos de preços. Este tipo de modelo trabalha com várias qualidades agregadas em uma única variável. XU et al. (1993), que desenvolveram um estudo para avaliar propriedades rurais utilizando modelos hedônicos, atribuíram este tipo de inovação aos avanços da Análise Multivariada e da Teoria Econômica. Segundo eles, “... *Quando aplicadas ao mercado de terra, as técnicas hedônicas assumem que uma dada parcela pode ser identificada por uma única classe de níveis de atributos, e o valor da parcela de*

terra é uma agregação dos valores de seus atributos individuais”. Vale ressaltar que a influência de cada atributo também pode ser identificada individualmente, conforme pôde ser observado em PASHA & BUTT (1996) em modelo de avaliação residencial desenvolvido para o Paquistão. No trabalho citado, os números de banheiros, quartos etc. foram agregados na variável número de cômodos e, após ajuste do modelo, a influência de cada atributo, separadamente, pode ser estimada.

Em 1977, um estudo elaborado por OTTENS MANN verificou que a taxa de crescimento da população nos Estados Unidos poderia ser uma boa variável explicativa para a valorização da terra. Em seu modelo, ele propôs que, além da taxa de crescimento populacional, a população e a renda média também fossem incluídas. Os resultados da regressão entre o preço e as três variáveis citadas foram bastante satisfatórios.

AZAR et al. (1994) propuseram um modelo para identificar as terras mais valorizadas em Beirute, através do uso de SIG. Em seu estudo, eles utilizaram quatro fatores de valorização: declividade, vista para o mar, vista para a cidade e distância às rodovias principais. Na impossibilidade de determinar qual é mais importante que outra, todas as variáveis foram consideradas como tendo pesos iguais. Os valores da terra foram, então, divididos em três classes, de acordo com o seu padrão, sendo os da classe 1 (declividade $< 15^\circ$, alta visibilidade para o mar e para a cidade, e distância de 200m ou menos até as rodovias), os mais valorizados.

ABELSON (1997) estudou um modelo teórico sobre valorização de terrenos e casas e através de deduções matemáticas, concluiu que a distância à zona central de negócios (Central Business District – CBD) explica uma grande porcentagem da variação dos preços. Na parte prática deste trabalho, ele estuda o comportamento dos preços imobiliários entre 1931 e 1989. Dentre as conclusões obtidas, salienta-se a diminuição do valor do m^2 de terrenos e residências, de acordo com a acessibilidade em épocas diferentes e a verificação de que, ao longo de todo este tempo, os preços caíram, geralmente exponencialmente, com a distância ao CBD. De maneira geral, a distância ao CBD explicou 50% das diferenças no valor da terra em 1931 e 1948, sendo que em 1968 esta variável explicou 28% das diferenças em preços. A queda nos preços de terrenos e residências, de acordo com

a distância ao CBD, entre os anos 1931 e 1968 é mostrada na Figura 2.1. Para o primeiro período da década de 70, o autor verificou que 80% das variações em preços de casas foram explicadas pela distância ao CBD e custo de transporte, tamanho e qualidade do ambiente local. Entre 1977 e 1989, preços de residências tiveram uma forte relação com o acesso ao CBD, não havendo evidência de que melhoramentos ambientais causaram alteração significativa em termos de valor. Para a modelagem de preços de terrenos, as seguintes variáveis foram consideradas: distância ao CBD, melhoramentos das estradas (variável binária, sendo o valor 1 atribuído a melhoramentos significantes e 2 para grandes melhorias) e proporção de propriedades recebendo serviços de esgoto.

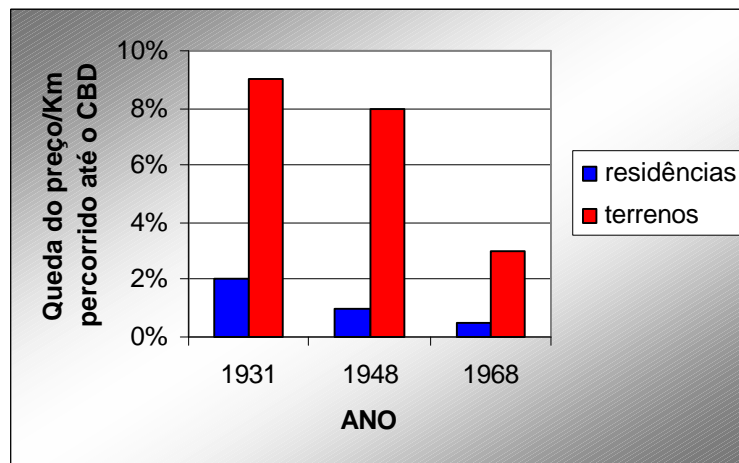


Figura 2.1 – Queda nos preços de residências e terrenos obtida pelo estudo de ABELSON (1997).

No estudo feito por CHESHIRE & SHEPPARD (1995) desenvolvido para verificar a valorização de imóveis nos EUA, observou-se que o preço da propriedade não varia apenas com a acessibilidade mas, também, com características da vizinhança, serviços públicos locais e outras características não físicas das residências. A avaliação de residências desenvolvida pelos autores mostrou que o modelo de avaliação melhora substancialmente quando características espaciais, como por exemplo, existência de escolas nas proximidades, são adicionadas.

LEVIN & WRIGHT (1997) utilizam séries temporais para explicar a flutuação de preços em Londres e no Reino Unido no período de 1965 a 1995.

Dentre as relações observadas por eles, salienta-se uma forte relação entre preços e renda familiar, sendo que em épocas de financiamentos e de mais alta renda, as pessoas se dispuseram a pagar mais pelo mesmo bem. Outro fato interessante observado é que existe uma relação positiva entre o crescimento dos preços imobiliários no passado e no futuro, principalmente para Londres. De maneira geral, a análise desenvolvida apresentou indícios de que a especulação é, possivelmente, um fator determinante em preços de residências. Além destes autores, pode-se citar o trabalho de GIBB (1994), no qual foram desenvolvidos modelos para estimar preços de residências em Glasgow e Edinburgh.

FORREST et al. (1996) também estudaram a influência da acessibilidade nos valores residenciais na grande Manchester, situada no norte da Inglaterra. Em seu trabalho, foi estudada a influência da construção de uma linha de metrô no valor das propriedades da região. Ao contrário do que imaginavam, foi constatado que propriedades muito próximas à nova linha desvalorizam, principalmente devido à perturbações, barulhos etc. De forma geral, a nova linha não trouxe grandes modificações nos valores residenciais situados ao longo da mesma.

O estudo de caso apresentado por PHANG & WONG (1997) para Singapura mostra que as taxas de juros, crescimento de renda e oferta de habitações não tiveram influência significativa no valor de propriedades privadas. No estudo em questão, eles observaram que os preços de imóveis privados estavam fortemente correlacionados com os preços de imóveis públicos. Vale ressaltar que o setor público exerce uma forte influência sobre o sistema habitacional local. Somente para exemplificar, em 1995, 86% do conjunto de imóveis residenciais do país era composto por casas populares.

As Redes Neurais se mostraram bastante eficientes em estudo realizado por ALMOND et al. (1997). Em seu trabalho, algumas aproximações foram consideradas, utilizando uma nova forma de aprendizado, onde, inicialmente uma rede não supervisionada é usada para estabelecer classes, que servirão como entrada para as redes convencionais. Este critério, segundo os autores, provocou uma melhora de 10% em relação às previsões fornecidas pelo método convencional.

Pode-se citar ainda um modelo proposto por AOKI et al. (1994), que consideraram o conceito de custos de transportes, que foi definido não só como carga monetária mas também como *stress* mental e físico sofrido dentro dos meios de transporte de uma cidade. Estes custos foram definidos em termos de acessibilidade. O primeiro modelo proposto, que só levou em consideração a acessibilidade, forneceu um coeficiente de determinação de 0,79, mas não obedeceu outras suposições do modelo de regressão. O segundo modelo incluiu as atividades urbanas do local: zona de uso da terra (distrito comercial, residencial etc., num total de 9 categorias), densidade (num total de 14 categorias) e tipo de uso da terra atual (floresta, rios, etc., num total de 16 categorias). Este último proporcionou uma melhora substancial do coeficiente de correlação (0,90) e variância dos resíduos constante. Como resultado, chegou-se à conclusão de que somente os custos de transportes não são suficientes para estimar os valores das propriedades.

De maneira geral, pode-se notar que o emprego de alguma medida de acessibilidade (geralmente a distância ao distrito central de negócios) é uma constante nos modelos citados. Ela pode servir como uma medida do espalhamento urbano, que cada vez mais tem preocupado avaliadores e planejadores, não só pelos custos adicionais que ele proporciona à sociedade, como também pela má qualidade de vida, segregação social e problemas ecológicos que ele pode representar. Neste contexto, o trabalho de PRATT (1996) faz uma crítica ao problema sob vários aspectos. Do ponto de vista ecológico, tem-se que, cidades espalhadas tendem a usar mais intensamente os meios de transporte, com conseqüente aumento de poluição. A piora na qualidade de vida se dá em função do maior tempo despendido para viajar de um local a outro, tempo este que poderia ser gasto com descanso ou lazer. Como se não bastasse, a estrutura dos meios de transporte permite que alguns grupos sejam melhor atendidos, dificultando o acesso a outros grupos menos privilegiados. Pior que isso, geralmente o transporte é mais acessível a pessoas de mais alta renda. Se o problema não apresenta solução tão imediata, uma forma de amenizá-lo seria o emprego de alguma ferramenta eficiente, que seja capaz de extinguir, ou caso isso não seja possível, reduzir bastante os vazios urbanos. O primeiro passo para isso poderia ser a implantação de uma política de impostos, que

desmotivasse os proprietários deste tipo de imóvel a possuí-los. O autor afirma que, para tal fim, um critério de avaliação adequado torna-se imprescindível, uma vez que a taxa sobre a propriedade deve ser baseada em seu valor monetário.

2.6. PRÓS E CONTRAS DOS IMPOSTOS SOBRE A PROPRIEDADE

Assim como ocorre em várias cidades brasileiras, existem cidades e até mesmo países, que, de certa forma, têm sua sobrevivência econômica garantida através dos impostos sobre a propriedade. Dentre estes países, pode-se citar o Reino Unido, onde a maior parte da receita do Estado provém deste tipo de imposto. Portanto, uma forma correta de avaliação imobiliária torna-se imprescindível, uma vez que as taxas sobre a propriedade são calculadas a partir de seu valor.

O que diferencia os bens imóveis é sua fixação no espaço, alto valor unitário e a longa vida útil. Sendo assim, torna-se mais simples implantar e vigiar impostos sobre a propriedade. Além disso, o sistema de taxaço sobre o valor da terra, *a priori*, pode ser pensado como ferramenta potencial para resolver problemas como espalhamento urbano, concentração de renda, falta de investimentos, especulação imobiliária etc. WYATT (1994) contesta esta teoria, assegurando que mantendo o sistema atual, incentivos como abatimentos ou incrementos de taxas são fatores pouco importantes para atingir os padrões desejados de uso da terra e desenvolvimento econômico. Ele justifica seu ponto de vista afirmando que o incentivo dado através de redução de taxas para construções, embora traga um pequeno efeito, não garante que áreas que não precisem de estímulo também atraiam investimento, muitas vezes até maiores que áreas carentes. Quanto à distribuição de renda, o autor afirma que, quanto maiores as taxas cobradas, maior a desigualdade social. Diante destes fatores, ele defende a implantação de um sistema de imposto progressivo como única solução viável para os problemas mencionados anteriormente.

Outro trabalho que coloca em questão a importância do uso de taxas para baratear o preço da terra urbana é apresentado por ROAKES (1996). Em seu

artigo, ela contesta o uso da taxa por si só, apregoando que as taxas sobre a propriedade devem ser separadas das taxas de melhorias.

Se por um lado, autores contestam o uso de impostos sobre a propriedade como ferramenta de planejamento urbano, por outro, alguns deles conseguem mostrar que a influência causada por tais taxas não é tão insignificante assim. Este é o caso de LIM (1992), que usou um modelo teórico para verificar a influência destas taxas sobre a estrutura urbana. Ele constatou que terrenos vagos, próximos a zonas residenciais, se sofrerem majoração de tributos, tendem a ser utilizados para fins residenciais.

O trabalho de BENTICK (1996) se situa entre os dois extremos, afirmando que “...as taxas sobre a propriedade urbana precisam ser suplementadas por controles de planejamento para prevenir que consumidores na periferia urbana se mudem para a área rural para fugir das taxas”. Em seu estudo, ele concluiu que para uma área não desenvolvida, uma taxa que não discrimine a terra urbana da rural, não afeta a periferia urbana, uma vez que as pessoas não sentem necessidade de se mudar para a área rural a fim de fugir dos impostos (caso comum nos países com forte dependência do meio de transporte automóvel). Por outro lado, uma taxa discriminatória irá contrair a periferia urbana, mas, também reduzirá a população urbana, exceto se ela for reforçada com regulamentos para zoneamento que previnam a mudança de pessoas para o outro lado da fronteira.

3. REDES NEURAIIS, REGRESSÃO LINEAR MÚLTIPLA E SIG

A primeira seção deste capítulo consta de uma introdução sobre a Inteligência Artificial, onde um breve histórico sobre as Redes Neurais, com os principais nomes envolvidos no seu desenvolvimento, será traçado. Na seqüência, o modelo de Rede Neural a ser utilizado para o desenvolvimento deste trabalho, ou seja, a Rede Multilayer Perceptron, será analisado em maiores detalhes. A próxima etapa faz uma revisão sobre a Regressão Múltipla, que é o método utilizado para avaliar o desempenho das Redes Neurais. Finalizando o capítulo, outra ferramenta amplamente utilizada neste trabalho será apresentada: o Sistema de Informação Geográfica (SIG).

3.1 ASPECTOS DA INTELIGÊNCIA ARTIFICIAL

O desenvolvimento de sistemas inteligentes partiu da sede do ser humano por conhecimento e do desejo de construir máquinas que executem certas tarefas tão bem quanto ele.

A arquitetura utilizada pela maioria dos computadores convencionais compreende uma simples unidade de processamento central conectada a uma área de memória. O processador então executa, de forma seqüencial, um programa

armazenado na memória. Portanto, este tipo de computador apenas consegue executar as tarefas para as quais ele foi programado.

Por outro lado, os neurocomputadores buscam modelar a estrutura do cérebro humano e a maneira pela qual ele consegue processar informação. Para tal intuito, é necessária uma nova arquitetura computacional, que, assim como o cérebro humano, seja composta de um grande número de elementos de processamento altamente conectados, que operem de uma maneira paralela.

O cérebro humano consegue executar certas atividades muito mais rapidamente que o computador, enquanto que outras não. Por exemplo, o computador executa cálculos numéricos a uma velocidade muito superior ao cérebro, enquanto que, para a máquina, o reconhecimento de uma imagem ou de um som é muito mais lento e difícil.

HAYKIN (1994) define o cérebro humano como um computador com estrutura altamente complexa, não linear e paralela que tem a capacidade de organizar neurônios para executar certos tipos de tarefas. É este paralelismo do cérebro que garante a ele a capacidade de armazenar e representar o conhecimento adquirido e torná-lo bastante acessível. O paralelismo também o torna capaz de processar este conhecimento simultaneamente (BEALE e JACKSON, 1990). A forma como o cérebro consegue executar certas tarefas é uma incógnita, tornado difícil a sua sintetização. Porém, sabe-se que ele é formado de muitas unidades de processamento altamente interconectadas, os neurônios. Cada neurônio é ligado a outros milhares. E é através do conhecimento do funcionamento destas unidades que é possível, hoje em dia, simular, ainda que precariamente, o comportamento cerebral.

3.1.1 ANALOGIA BIOLÓGICA

O cérebro (humano ou animal) é composto de células chamadas neurônios. Estima-se que um cérebro humano possua mais de 100 bilhões de neurônios, e mais de uma centena de tipos destes. Os neurônios funcionam em grupos, denominados redes, sendo que cada rede pode ser formada por milhares de

neurônios interconectados. Portanto, o cérebro pode ser visto como uma coleção de Redes Neurais. Tanto o pensamento quanto o comportamento inteligente são controlados pelo sistema nervoso central e pelo cérebro (TRIPPI e TURBAN, 1992).

Cada neurônio é formado de dendritos, núcleo e axônio. Ele é uma estrutura simples, que permite três funções básicas: entrada, processamento e saída de sinais. Os dendritos são conexões através das quais os sinais de entrada chegam aos neurônios. O núcleo é o processador do neurônio. Ele recebe os sinais de entrada através dos dendritos e os adiciona de alguma forma. Se o valor resultante estiver acima de um certo limite, o neurônio se excita e tende a propagar o estímulo. Caso contrário, ele fica inibido. Os axônios servem como o canal de saída do neurônio. Eles estão ligados aos dendritos de outros neurônios através das sinapses. As sinapses não são ligações físicas, mas sim químicas e temporárias, e se dão através de neurotransmissores.

3.1.2. HISTÓRICO – REDES NEURAIIS

Em 1943, McCulloch e Pitts propuseram um modelo de neurônio artificial, como uma unidade de processamento simples, cuja proposta não era copiar o neurônio verdadeiro, mas, sim, modelá-lo (BEALE e JACKSON, 1990). A rede empregada por eles empregava um modelo extremamente simples de neurônio, cada um tendo tanto entradas excitatórias quanto inibitórias. Quando o número de entradas excitatórias excedesse o número de entradas inibitórias, o neurônio sofria uma ativação e produzia uma saída igual a 1 (LUGER e STUBBLEFIELD, 1989). O neurônio proposto não apresentava *threshold* (limiar), que corresponde a um valor que determina se o neurônio vai responder ou não a determinada entrada. Se a entrada total for menor que este limiar, o neurônio permanece inibido. Caso contrário, ele se excita e tende a propagar o estímulo.

Alguns anos depois, Von Neumann mostrou que, quando redundâncias eram introduzidas nas redes McCulloch-Pitts, elas poderiam continuar funcionando normalmente apesar do mal funcionamento de elementos individuais.

Este foi um resultado importante porque faz uma sugestão de como o cérebro humano continua funcionando apesar da perda de células nervosas ao longo de nossa vida (LUGER e STUBBLEFIELD, 1989).

Em 1949, Hebb propôs uma forma de aprendizado para as Redes Neurais, à qual ele denominou aprendizado *hebbiano*. Por esta lei de aprendizado, se dois neurônios estão ativos simultaneamente, a conexão entre eles deve ser reforçada (HAYKIN, 1994).

Nos anos 50, Rosenblatt adicionou *thresholds* aos neurônios e propôs um procedimento pelo qual a rede poderia aprender, ajustando sistematicamente seus pesos sinápticos. Esta rede foi batizada de *perceptron* e podia aprender problemas simples de classificação (LUGER e STUBBLEFIELD, 1989). Porém, este modelo era bastante limitado. Em 1969, Minsky e Papert mostraram que a rede *perceptron* não era capaz de resolver o problema do ou-exclusivo, ou seja, ela só poderia classificar padrões que eram linearmente separáveis.

A partir destas descobertas, houve um desaquecimento no estudo sobre Redes Neurais, que só voltaram à tona nos anos 80, quando alguns problemas apresentados pelo *perceptron* passaram a ser resolvidos através do uso de novos modelos. Dentre os novos modelos, pode-se citar as redes de Hopfield (BEALE & JACKSON, 1990), a máquina de Boltzmann (HAYKIN, 1994) e as redes com várias camadas.

3.1.3. REDES NEURAIIS ARTIFICIAIS

As Redes Neurais artificiais são baseadas em modelos abstratos do funcionamento do cérebro humano e tentam reproduzir sistemas biologicamente realísticos.

Dentre os benefícios da computação neural, pode-se citar (TUBB, 1993):

- **habilidade de aprender com exemplos:** os computadores neurais têm a capacidade de aprender com a experiência, objetivando

melhorar seu desempenho e se adaptar a ambientes novos e dinâmicos, diferentemente dos computadores comuns;

- **robustez:** as Redes Neurais tem habilidade em lidar com ruídos. Elas são tolerantes à falhas e podem apresentar degradação gradual, ou seja, apesar de alguma falha no sistema, elas continuam a fornecer respostas adequadas por um bom tempo, o que as diferencia dos computadores convencionais, onde uma falha pode causar prejuízo do sistema como um todo;
- **velocidade de processamento:** como as Redes Neurais consistem de um grande número de unidades de processamento operando em paralelo, elas podem operar em velocidades consideráveis em relação aos métodos computacionais comuns.

Além das vantagens já mencionadas, em muitos casos e dependendo do problema para o qual são utilizadas, o desempenho das Redes Neurais têm se mostrado bastante superior aos métodos estatísticos usados para o mesmo fim (FALAS, 1995). SUBRAMANIAN et al. (1993), fizeram um estudo onde compararam as Redes Neurais com alguns métodos estatísticos de classificação. A conclusão a que chegaram é que, em todas as circunstâncias, incluindo tamanhos de amostras pequenos e funções mais complexas, as Redes Neurais apresentaram soluções bastante superiores.

Um outro estudo, feito por KWON et al. (1995), usou modelos de regressão múltipla e Redes Neurais para encontrar soluções para o problema do caixeiro viajante. Numa comparação dos dois métodos, verificou que as rotas fornecidas pelas Redes Neurais apresentaram superioridade. Ainda comparando regressão e Redes Neurais, pode-se citar o trabalho de GUEDES (1994), que usou modelos de regressão e Redes Neurais para fazer estimativas dos valores de apartamentos no Rio de Janeiro, também obtendo melhores resultados com a aplicação deste último.

3.1.3.1. O NEURÔNIO ARTIFICIAL

O neurônio artificial é uma unidade de processamento matematicamente simples, que recebe uma ou mais entradas e as transforma em saídas. Cada entrada tem um peso associado, que determina sua intensidade (TUBB, 1993). A Figura 3.1, a seguir, mostra um esquema de neurônio artificial.

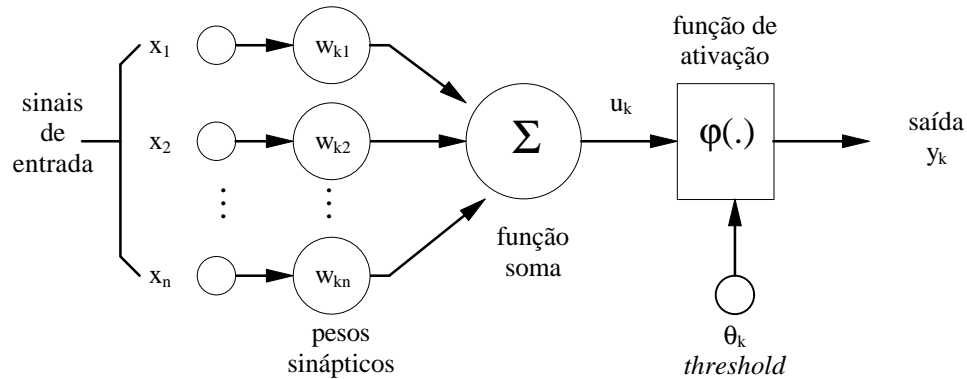


Figura 3.1 - Modelo não linear de um neurônio (HAYKIN, 1994).

É possível distinguir alguns elementos importantes no neurônio mostrado na Figura 1:

1. as **sinapses**, que são caracterizadas por um peso, w , que pode representar a sua intensidade. O papel do peso w_{kj} é multiplicar o sinal x_j na entrada da sinapse j , conectada a um neurônio k . O peso w_{kj} é positivo se a sinapse associada é excitatória e negativo se a sinapse associada é inibitória;
2. um **somatório**, que adiciona as entradas ponderadas pelos seus pesos respectivos, ou seja,

$$u_k = \sum_{i=1}^n w_{ij} x_j \quad (1)$$

3. um **limiar** (*threshold*), θ_k , que tem um papel determinante na saída do neurônio. Se o valor de u_k for menor que este limiar, então, a saída do neurônio fica inibida. Caso contrário, o neurônio fica ativo.
4. uma **função de ativação**, que funciona como um limitante à amplitude da saída do neurônio, ou seja, a entrada é normalizada dentro de um intervalo fechado, comumente $[0,1]$ ou $[-1,1]$;
5. a **saída** do neurônio, y_k , onde:

$$y_k = \mathbf{j} (u_k - \mathbf{q}_k) \quad (2)$$

onde φ é a função de ativação.

Em geral, o valor do *threshold* é aplicado com a inclusão de uma entrada x_0 igual a -1 e um peso w_{k0} igual ao valor de θ_k . Portanto, a nova entrada da função de ativação, já incluindo o limiar, é dada por:

$$v_k = \sum_{j=1}^n w_{kj} x_j - \mathbf{q}_k \quad (3)$$

3.1.3.2. TIPOS DE FUNÇÃO DE ATIVAÇÃO

A função de ativação define a saída do neurônio em termos do nível de atividade do mesmo (HAYKIN, 1994). Dentre as funções de ativação mais comuns, se enquadram as seguintes:

- **função linear:** $\varphi_i(t+1) = v_i(t)$
- **função threshold ou limiar:** $\varphi_i(t+1) = \begin{cases} 1, & \text{se } v_i(t) \geq \theta \\ 0, & \text{se } v_i(t) < \theta \end{cases}$
- **função sigmóide logística:** $\varphi_i(t+1) = 1/(1 + e^{-v_i(t)})$
- **função tangente hiperbólica:** $\varphi_i(t+1) = (1 - e^{-v_i(t)}) / (1 + e^{-v_i(t)})$

- **função linear por partes:** $\varphi_i(t+1) = \begin{cases} +1, & \text{se } v_i(t) > \theta \\ -1, & \text{se } v_i(t) < \theta \\ v_i(t), & \text{em outra parte} \end{cases}$

3.1.3.3. TOPOLOGIA DA REDE

A maneira como os neurônios são organizados é chamada de topologia da rede. A topologia irá afetar o desempenho da rede, assim como as aplicações para as quais ela é desejada, e sua estrutura está intimamente ligada ao algoritmo de aprendizado usado para a fase de treinamento. Algumas redes permitem que as conexões caminhem tanto no sentido entrada-saída, quanto saída-entrada. Outras permitem que os neurônios da mesma camada estejam conectados. Ainda há as que permitem que o neurônio envie sinais de volta para ele mesmo (TUBB, 1993).

Dentre as topologias, pode-se citar a *Multilayer Perceptron*, que é formada de uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída e é o tipo de arquitetura mais encontrado. Há também a *rede de Kohonen*, que é formada por uma camada de entrada e uma única camada de saída, onde cada neurônio está conectado a todos os seus vizinhos. Na *rede de Hopfield*, não há neurônios de entrada ou saída. A entrada de um neurônio é a saída dos demais neurônios e a sua saída vai ser a entrada do demais neurônios da rede. A rede *ART* é formada por camada de entrada e saída, além de controladores de fluxo de sinais. Para este estudo foi escolhida a rede *Multilayer Perceptron* devido ao fato desta ser mais adequada para fazer classificações deste tipo, além de ser a mais conhecida e utilizada, tendo, inclusive, *softwares* para sua implementação disponíveis na Internet.

3.1.3.4. ALGORITMOS DE APRENDIZADO

Segundo CARVALHO et al. (1998), denomina-se algoritmo de aprendizado a um conjunto de regras bem definidas para a solução de um problema

de aprendizado. Existem vários tipos de algoritmos de aprendizado, que são específicos para determinados modelos de redes neurais e que diferem entre si, principalmente, pela forma com que os pesos são atualizados.

Os paradigmas de aprendizado, por sua vez, definem a maneira como a rede se relaciona com o ambiente e se dividem em três grupos principais: supervisionado, não supervisionado e híbrido.

Para o aprendizado supervisionado, apresenta-se à rede, na fase de treinamento, um conjunto de entradas acompanhadas de suas respectivas saídas. O objetivo é minimizar o sinal de erro, que é uma função da diferença entre a saída desejada e aquela fornecida pela rede. Esta minimização se dá pelo ajuste dos pesos da rede. Um exemplo deste paradigma de aprendizado é o método *backpropagation*, que será discutido mais adiante.

No aprendizado não supervisionado, a rede aprende sozinha, sem uma mensagem de erro externa, ou seja, não há supervisão externa. É necessário que entradas parecidas sejam apresentadas à rede, para que esta possa extrair características estatisticamente relevantes e criar classes de maneira automática.

O aprendizado híbrido, por sua vez, consiste de uma combinação dos aprendizados supervisionado e não supervisionado. Um exemplo é o aprendizado por reforço, onde a rede aprende de seu próprio ambiente, a partir dos dados de entrada. A única informação externa que a rede recebe é a indicação de que a resposta fornecida está correta ou não.

3.1.3.5. A REDE *Multilayer Perceptron*

Como mencionado anteriormente, a rede *perceptron*, que é constituída de apenas duas camadas (uma de entrada e outra de saída) só consegue classificar padrões que sejam linearmente separáveis. Diante deste problema, uma possível aproximação seria usar mais que um *perceptron*, cada um sendo empregado para distinguir seções linearmente separáveis da entrada e então combinando suas saídas em outro *perceptron*, que poderia produzir uma indicação

final da classe a que aquela entrada pertence (BEALE e JACKSON, 1990). A Figura 3.2 mostra esta idéia graficamente.

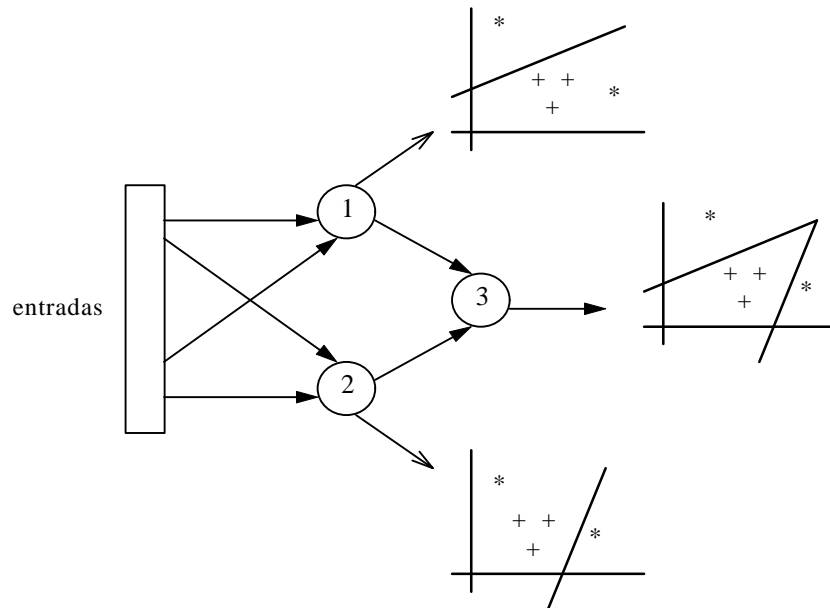


Figura 3.2 - Comportamento de classificação de dois padrões de uma rede com uma camada intermediária.

A nova rede arranja as unidades *perceptron* em camadas e o novo modelo passa a ser chamado *multilayer perceptron*. O novo modelo de rede passa a contar com uma ou mais camadas de neurônios intermediárias. Cada unidade pertencente à camada intermediária funciona como o *perceptron*, exceto porque sua função de ativação não é mais a função degrau e sim, a função sigmóide ou a função *threshold* linear. Esta escolha pelo tipo de função de ativação se dá porque, com a utilização de uma das duas alternativas apresentadas, a atualização dos pesos se dá de forma mais suavizada, ou seja, os pesos não vão sofrer alterações bruscas. A topologia de uma rede *multilayer perceptron* com três camadas (uma intermediária) é mostrada na Figura 3.3.

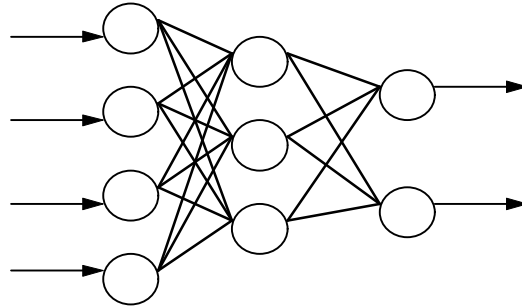


Figura 3.3 - Topologia de uma rede *multilayer perceptron* com uma camada intermediária.

3.1.3.5.1. ALGORITMO DE APRENDIZADO PARA UMA REDE MLP

A regra de aprendizado utilizada para treinar uma rede *multilayer perceptron* é denominada Regra Delta Generalizada ou, mais comumente, *backpropagation*.

Quando um padrão é apresentado à rede pela primeira vez, esta produz uma saída aleatória. A diferença entre esta saída e a desejada constitui o erro. A intenção do trabalho de treinamento é buscar cada vez mais diminuir o valor deste erro. Para tal intuito, o valor dos pesos deve ser ajustado a cada nova iteração. A regra *backpropagation* proporciona que os pesos da camada de saída sejam os primeiros a serem ajustados e, posteriormente, os pesos das demais camadas, de trás para frente (daí o nome da regra).

Seja E_p , a função erro para o padrão p , d_{pj} , o valor da saída desejado (para o padrão p e o nó j), o_{pj} , o valor da saída obtido e w_{ij} , o peso do nó i para o nó j . Então, a função E_p pode ser dada por:

$$E_p = \frac{1}{2} \sum_j (d_{pj} - o_{pj})^2 \quad (4)$$

Portanto, o objetivo é minimizar esta função de erro. Se esta função for simples, a rede tem garantia de encontrar solução. Caso contrário, a rede pode cair em um mínimo local e não produzir uma saída satisfatória.

A atualização dos pesos se dá pela adição de um termo de variação a eles (Δw_{ij}), ou seja, para um peso w_{ij} , o seu peso atualizado vai ser dado por:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \Delta w_{ij} \quad (5)$$

onde :

$$\Delta w_{ij} = \eta x_i o_j (1 - o_j) d_j \quad (6)$$

$$d_j = (d_j - o_j), \text{ para última camada} \quad (7)$$

$$d_j = \sum w_{jk} d_k, \text{ para camadas intermediárias} \quad (8)$$

A variável η é denominada taxa de aprendizado. Seu valor vai determinar o quão suavemente se dará a atualização dos pesos. Se η for grande, o valor de Δw_{ij} será grande e, conseqüentemente, os pesos sofrerão uma alteração mais brusca.

Uma outra variável, que pode ser incluída na atualização de pesos, é o *momentum* (α). Este novo termo pode aumentar a velocidade do aprendizado e tem por característica acelerar o treinamento em regiões planas da superfície de erro. Com a adição do *momentum*, Δw_{ij} passa a ser dado por:

$$\Delta w_{ij}(t+1) = \eta x_i o_j (1 - o_j) d_j + \alpha (w_{ij}(t) - w_{ij}(t-1)) \quad (9)$$

A atualização de pesos pode se dar de duas maneiras: por padrão ou por ciclo. Na primeira, os pesos são atualizados após a apresentação de cada padrão. Na segunda, os pesos são atualizados após a apresentação de todos os padrões. A maneira a ser utilizada depende da aplicação. Se o conjunto de dados é grande e redundante (poucos padrões e muitas entradas, por exemplo), a primeira é melhor. Caso contrário, quando o conjunto apresenta muitos padrões a serem classificados, é melhor utilizar a atualização por ciclo.

O algoritmo *backpropagation* apresenta alguns problemas, quanto às dificuldades de aprendizado. Em superfícies muito complexas, o treinamento é

muito lento e existe o risco da rede cair num mínimo local. Além disso, pode ocorrer *overfitting*, fenômeno que surge depois de um certo tempo de treinamento, quando a rede memoriza os padrões que já foram apresentados e o processo de classificação piora sensivelmente. Possíveis soluções para estes problemas são: diminuir a taxa de aprendizado, adicionar nós na camada intermediária, utilizar *momentum* e adicionar ruídos.

Finalmente, as Figuras 3.4 e 3.5 mostram diagramas esquemáticos dos algoritmos de treinamento e teste de uma rede *Multilayer Perceptron* (MLP).

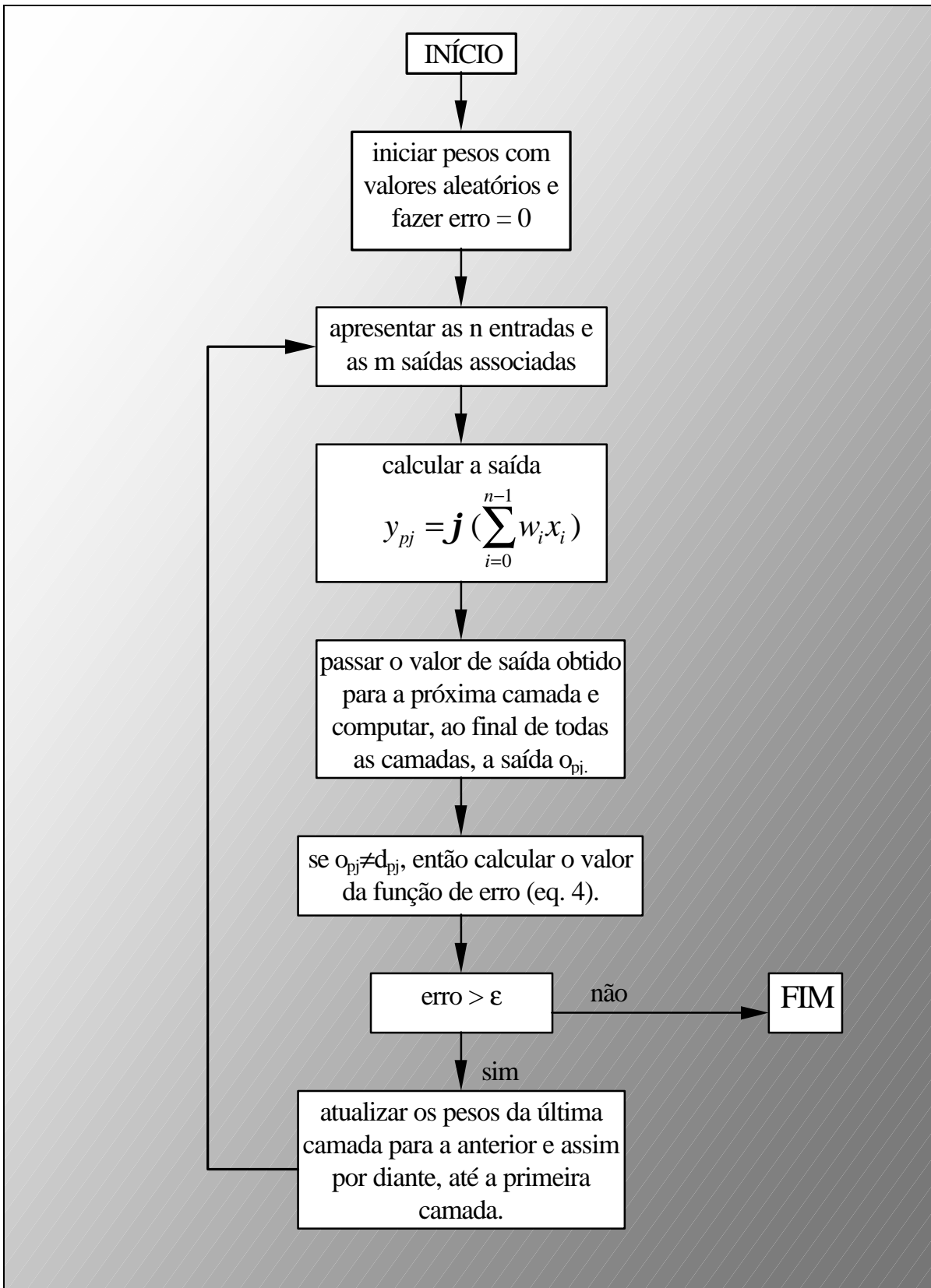


Figura 3.4 - Diagrama esquemático do treinamento de uma rede MLP.

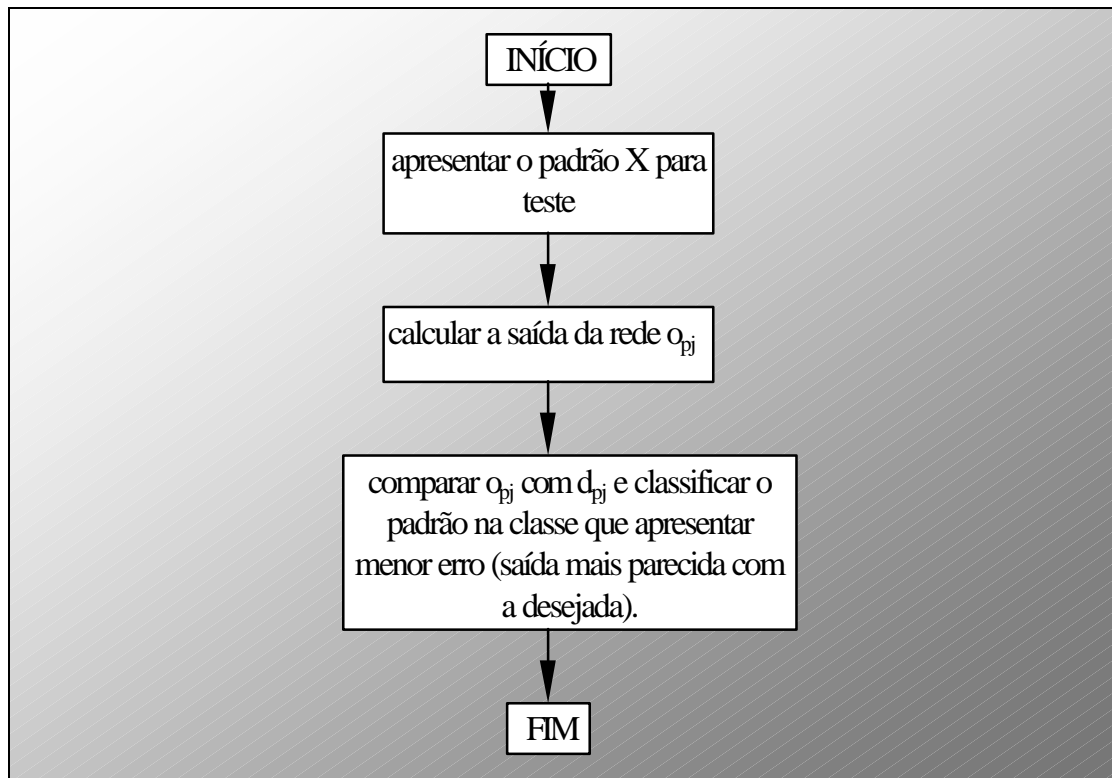


Figura 3.5 - Diagrama esquemático da fase de teste de uma rede MLP.

3.2. ANÁLISE DE REGRESSÃO

3.2.1. O MODELO LINEAR

O modelo de Regressão Linear Múltipla é usado para estudar a relação entre uma variável dependente e várias variáveis independentes. Sua fórmula genérica é dada por:

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{iK}) + e_i = \mathbf{b}_1 x_{i1} + \mathbf{b}_2 x_{i2} + \dots + \mathbf{b}_K x_{iK} + e_i, \quad i = 1, 2, \dots, n \quad (10)$$

onde:

y é a variável dependente ou explicada

x_1, x_2, \dots, x_K são as variáveis independentes ou explicativas

e o termo e_i representa o distúrbio aleatório, que pode se originar devido a diversas circunstâncias, tais como a ausência de fatores influentes no modelo e distúrbios causados por erros de medida.

3.2.1.1. SUPOSIÇÕES DO MODELO DE REGRESSÃO LINEAR CLÁSSICO

O Modelo de Regressão Linear Clássico consiste de uma série de suposições que descrevem as formas do modelo e também ditam os procedimentos adequados de estimação e inferência.

3.2.1.1.1. LINEARIDADE DO MODELO DE REGRESSÃO

Assuma que o vetor coluna \mathbf{x}_k representante as n observações na variável x_k , $k=1, 2, \dots, K$ e que estes dados estão alocados em uma matriz \mathbf{X} de dimensão $n \times K$. Na maioria dos casos, a primeira coluna da matriz \mathbf{X} é composta somente de 1's, de forma que \mathbf{b}_1 é o termo constante no modelo. Assumindo também que \mathbf{y} represente as n observações y_1, y_2, \dots, y_n e que $\boldsymbol{\varepsilon}$ é um vetor coluna contendo os n distúrbios, o modelo dado em (10) pode ser rescrito como:

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \dots + \mathbf{x}_K\beta_K + \boldsymbol{\varepsilon} \quad (11)$$

Ou, similarmente, na forma da Suposição 1 dada por:

$$\text{Suposição 1: } \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (12)$$

A suposição de linearidade do modelo de regressão inclui o distúrbio aditivo e deve ser da forma descrita pela equação (11) ou na forma das variáveis originais ou para alguma transformação adequada destas. Portanto, no contexto da regressão, a linearidade se refere à forma pela qual os parâmetros e o distúrbio entram na equação e não necessariamente à relação entre as variáveis envolvidas.

Neste contexto, uma relação não linear da forma $y = Ax^b e^e$ pode ser linearizada através do uso dos logaritmos de x e y : $\ln y = \alpha + \beta \ln x + \varepsilon$.

3.2.1.1.2. POSTO COMPLETO

Esta suposição diz respeito aos regressores do modelo, onde deve-se assumir que não existe relação linear perfeita entre as variáveis independentes. Para que isto ocorra, é necessário que a matriz X tenha posto completo, ou seja, que as colunas desta matriz sejam linearmente independentes e que haja pelo menos K observações. Desta forma, a suposição 2 pode ser escrita como:

$$\text{Suposição 2: } X \text{ é uma matriz } n \times K \text{ com posto } K \quad (13)$$

3.2.1.1.3 VALOR ESPERADO DO DISTÚRBO

Deve-se assumir que o valor esperado do distúrbio seja zero para cada observação, ou seja:

$$E[\varepsilon_i/X] = 0 \quad (14)$$

Assim, para o grupo completo de observações, a suposição 3 pode ser escrita como:

$$\text{Suposição 3: } E[\mathbf{e} / X] = \begin{bmatrix} E[\mathbf{e}_1 / X] \\ E[\mathbf{e}_2 / X] \\ \vdots \\ E[\mathbf{e}_n / X] \end{bmatrix} = 0 \quad (15)$$

Em muitos casos, esta suposição não é restritiva. Porém, se o modelo original não contém o termo constante, supor $E[\varepsilon_i] = 0$ poderia não ser uma decisão correta. Por um lado, se $E[\varepsilon_i]$ puder ser expressa como uma função linear dos \mathbf{x}_i ,

então, uma transformação do modelo poderia produzir distúrbios com média zero. Por outro lado, uma média dos distúrbios diferente de zero poderia constituir uma parte substantiva da estrutura do modelo. Assim, como regra geral, modelos de regressão não devem ser especificados sem o termo constante, a menos que isto seja ditado pela teoria pertinente. Se houver razões para especificar que a média do distúrbio seja diferente de zero, isto deve ser construído na parte sistemática da regressão, deixando no distúrbio somente a parte desconhecida de ε .

3.2.1.1.4. DISTÚRBIOS ESFÉRICOS

A quarta suposição diz respeito à variância e covariância dos distúrbios:

$$\text{Var} [\varepsilon_i/\mathbf{X}] = \sigma^2, \quad \text{para todo } i = 1, \dots, n \text{ e}$$

$$\text{Cov} [\varepsilon_i, \varepsilon_j/\mathbf{X}] = 0, \quad \text{para todo } i \neq j$$

A variância constante é chamada homocedasticidade e a falta de correlação entre os distúrbios é chamada não-autocorrelação. A não-autocorrelação não implica que as observações y_i e y_j são não correlacionadas, mas sim que os desvios dos valores observados destas com relação aos valores esperados são não correlacionados. Estas duas suposições implicam que:

$$E[\mathbf{ee}' / X] = \begin{bmatrix} E[\mathbf{e}_1\mathbf{e}_1 / X] & E[\mathbf{e}_1\mathbf{e}_2 / X] & \cdots & E[\mathbf{e}_1\mathbf{e}_n / X] \\ E[\mathbf{e}_2\mathbf{e}_1 / X] & E[\mathbf{e}_2\mathbf{e}_2 / X] & \cdots & E[\mathbf{e}_2\mathbf{e}_n / X] \\ \vdots & \vdots & \ddots & \vdots \\ E[\mathbf{e}_n\mathbf{e}_1 / X] & E[\mathbf{e}_n\mathbf{e}_2 / X] & \cdots & E[\mathbf{e}_n\mathbf{e}_n / X] \end{bmatrix} = \begin{bmatrix} \mathbf{s}^2 & 0 & \cdots & 0 \\ 0 & \mathbf{s}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{s}^2 \end{bmatrix} \quad (16)$$

que pode ser resumida na seguinte suposição:

$$\text{Suposição 4: } E[\varepsilon\varepsilon'/\mathbf{X}] = \sigma^2\mathbf{I} \quad (17)$$

Os distúrbios que encontram as suposições de homocedasticidade e não-autocorrelação são chamados distúrbios esféricos.

3.2.1.1.5. REGRESSORES NÃO-ESTOCÁSTICOS

Em geral, assume-se que \mathbf{x}_i é não-estocástico, o que permite ao analista escolher os valores dos regressores e então observar os y_i , ao contrário do que ocorreria em uma situação experimental. Assim, a suposição 5, que consiste da especificação das variáveis independentes, fica da seguinte maneira:

Suposição 5: \mathbf{X} é uma matriz de constantes conhecida, com dimensão $n \times K$ (18)

3.2.1.1.6. NORMALIDADE

É conveniente assumir que os distúrbios são normalmente distribuídos, com média zero e variância constante. Esta suposição, que é útil principalmente na construção de testes estatísticos, implica que as observações em ε_i são estatisticamente independentes, é apresentada a seguir:

Suposição 6: $\varepsilon/\mathbf{X} \sim \mathbf{N}(0, \sigma^2 \mathbf{I})$ (19)

3.2.1.2. O MÉTODO DOS MÍNIMOS QUADRADOS

Sejam os parâmetros desconhecidos da relação $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$, os objetos de estimação. Então, se \mathbf{b} e e_i são as estimativas amostrais de $\boldsymbol{\beta}$ e ε_i , então, a regressão populacional é dada por $E[y_i/\mathbf{x}_i] = \mathbf{x}_i' \boldsymbol{\beta}$, enquanto que sua estimativa é denotada por:

$$\hat{y}_i = \mathbf{x}_i' \mathbf{b} \quad (20)$$

Sendo assim, o distúrbio associado com a i -ésima observação é dado por:

$$\varepsilon_i = y_i - \mathbf{x}_i' \boldsymbol{\beta} \quad (21)$$

Para um dado valor de \mathbf{b} , ε_i pode ser estimado através da seguinte expressão:

$$e_i = y_i - \mathbf{x}_i' \mathbf{b} \quad (22)$$

A partir das definições, obtém-se a relação dada por (23), que é mostrada graficamente na Figura 3.6.

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i = \mathbf{x}_i' \mathbf{b} + e_i \quad (23)$$

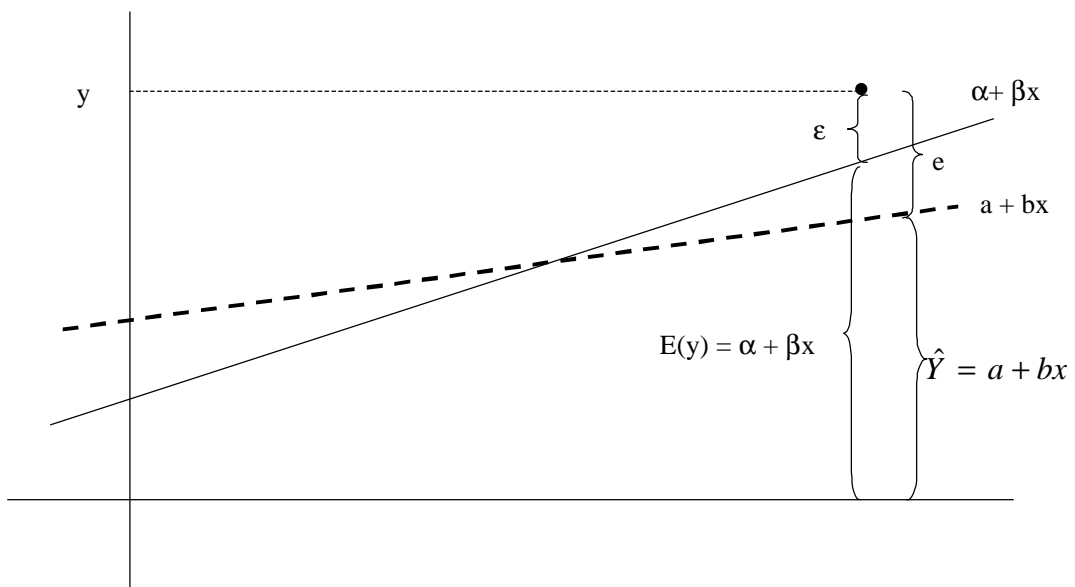


Figura 3.6 - Regressões amostral e populacional (GREENE, 1997).

3.2.1.2.1 O VETOR DOS COEFICIENTES OBTIDO POR MÍNIMOS QUADRADOS

O vetor dos coeficientes obtido por Mínimos Quadrados (β_0) deve ser escolhido de maneira a minimizar a Soma de Quadrados dos Resíduos, dada por:

$$\mathbf{e}_0' \mathbf{e}_0 = (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0) \quad (24)$$

Desta forma, para encontrar o valor de β_0 que minimiza a Soma de Quadrados dos Resíduos, deve-se diferenciar a expressão (24) com relação a β_0 e igualar o resultado a zero. Se a inversa de $(\mathbf{X}'\mathbf{X})$ existe, então a solução do sistema obtido, denotada por \mathbf{b} , é dada por:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (25)$$

3.2.1.2.1.1. TESTES DE HIPÓTESE E INTERVALOS DE CONFIANÇA PARA O COEFICIENTE β_K

Se o distúrbio aleatório ε possui Distribuição Normal Multivariada, então pode-se provar que a distribuição de \mathbf{b}/\mathbf{X} é dada por:

$$\mathbf{b}/\mathbf{X} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (26)$$

ou seja, \mathbf{b}/\mathbf{X} apresenta distribuição Normal Multivariada. Desta forma, cada elemento de \mathbf{b}/\mathbf{X} também é normalmente distribuído:

$$\mathbf{b}_k/\mathbf{X} \sim N(\beta_k, \sigma^2[(\mathbf{X}'\mathbf{X})_{kk}]^{-1}) \quad (27)$$

Usando estes resultados, pode-se elaborar testes de hipóteses ou construir intervalos de confiança para os elementos do vetor β . Para tanto, torna-se necessário conhecer a matriz de covariância

$$\text{Var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (28)$$

Se σ^2 é conhecido, uma estimativa da variância de b_k é dada pelo k -ésimo elemento da matriz dada em (28). Então, seguindo a suposição de normalidade, uma estatística de teste para β_k seria dada por:

$$z_k = \frac{b_k - \hat{a}_k}{\sqrt{\hat{\sigma}^2 S^{kk}}} \quad (29)$$

onde S^{kk} é o k -ésimo elemento da matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

Por outro lado, se σ^2 é desconhecido, então um estimador não viciado para a variância da regressão é dado por:

$$s^2 = \frac{e'e}{n - K} \quad (30)$$

A matriz de covariâncias estimada para \mathbf{b} pode então ser calculada através da seguinte expressão:

$$\text{Estimativa da Var}[\mathbf{b}] = s^2(\mathbf{X}'\mathbf{X})^{-1} \quad (31)$$

Conseqüentemente, a variância de b_k é dada pelo k -ésimo elemento da matriz dada em (31). Se s^2 for utilizado em lugar de σ^2 e se b e s^2 são independentes, pode-se mostrar que a razão:

$$t_k = \frac{b_k - \hat{b}_k}{\sqrt{s^2 S^{kk}}} \quad (32)$$

tem distribuição t-Student com $(n-K)$ graus de liberdade. Utilizando este mesmo resultado, pode-se construir intervalos de confiança para os β_k 's com nível $(1-\alpha)$ de confiança:

$$\text{Prob}(b_k - t_{\lambda/2} s_{bk} < \beta_k < b_k + t_{\lambda/2} s_{bk}) = 1 - \alpha \quad (33)$$

3.2.1.3. VERIFICAÇÃO DA ADEQUABILIDADE DA REGRESSÃO AJUSTADA

Uma medida de adequabilidade da regressão ajustada pode ser dada pelo coeficiente de determinação, R^2 . Este coeficiente serve para verificar o percentual da variância de y que é explicado pelo modelo ajustado e, para seu cálculo, alguns valores devem ser conhecidos *a priori*. O primeiro destes valores é a Variação Total em y (SST), cujo valor é obtido pela expressão (34).

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (34)$$

Se SSE denota a Soma de Quadrados dos Resíduos (e'e), então pode-se mostrar que:

$$SST = SSR + SSE \quad (35)$$

onde SSR: Soma de Quadrados devido à Regressão.

De posse destes valores, pode-se, então, calcular o coeficiente de determinação:

$$R^2 = \frac{SSR}{SST} \quad (36)$$

O valor do R^2 pode variar de 0, o que ocorre quando todos os elementos do vetor b , exceto o termo constante, são iguais a zero. Neste caso, os valores preditos para y são sempre iguais a \bar{y} e, portanto, diferentes valores para x não conduzem a diferenças nos valores preditos para y , ou seja, x não possui poder explicativo. Por outro lado, o valor máximo para o R^2 é igual a 1, o que ocorre quando os resíduos do modelo são iguais a zero e, portanto, 100% da variância de y é explicada pelo modelo (situação ideal). Portanto, quanto mais próximo de 1 estiver o valor do R^2 , mais adequado é o modelo de regressão ajustado.

Os cálculos necessários para a obtenção do R^2 podem ser resumidos em uma única tabela, denominada Tabela de Análise de Variância.

Tabela 3.1 - Tabela de Análise de Variância.

	Fonte	Graus de liberdade	Quadrado Médio
Regressão	$b'X'y - n\bar{y}^2$	$K - 1$	
Resíduo	$e'e$	$n - K$	s^2
Total	$y'y - n\bar{y}^2$	$n - 1$	$S_{yy}/(n-1) = s_y^2$
Coeficiente de determinação		$R^2 = 1 - e'e/(y'y - n\bar{y}^2)$	

Os resultados mostrados anteriormente também podem ser utilizados para testar a significância da regressão. Para tal, torna-se necessário testar a hipótese de que todos os coeficientes, exceto o termo constante, são iguais a zero. Conforme já foi mencionado, se todos os coeficientes são iguais a zero, então R^2 também é igual a zero e, portanto, o teste pode ser construído com base neste valor. Assim, se $\beta_2 = 0$ (onde β_2 é a parte do vetor β que não inclui a constante) e os distúrbios têm Distribuição Normal, pode-se mostrar que a estatística:

$$F = \frac{R^2 / (K - 1)}{(1 - R^2) / (n - K)} \quad (37)$$

tem Distribuição F com $(K-1)$ e $(n-K)$ graus de liberdade. Grandes valores para F são induzidos por altos valores do R^2 , o que contradiz a hipótese inicial.

Embora não se aplique a este estudo e, portanto, não será assunto incluído neste capítulo, pode-se generalizar o modelo clássico de regressão para situações onde o distúrbio não apresenta Distribuição Normal ou a matriz \mathbf{X} não é estocástica. Para maiores detalhes, ver GREENE (1997).

3.2.1.4. MULTICOLINEARIDADE

O problema da multicolinearidade existe quando as variáveis independentes são excessivamente autocorrelacionadas, dificultando uma análise precisa de seus efeitos individuais. Este tipo de problema fere uma das suposições básicas do modelo clássico de regressão que é a de que a matriz \mathbf{X} deve ter posto completo, ou seja, não deve existir relação linear perfeita entre as variáveis independentes. Esta suposição é feita para assegurar a existência da inversa de $\mathbf{X}'\mathbf{X}$.

São dois os possíveis casos de colinearidade: perfeita ou próximo da multicolinearidade. No primeiro caso, há um número infinito de vetores de parâmetros que podem ser consistentes com o mesmo valor esperado. Assim, independentemente de quantos dados pode-se obter, é impossível estimar os parâmetros do modelo. No segundo caso, o posto da matriz \mathbf{X} é $L < K$. Segue disso que na matriz \mathbf{X} há somente $L < K$ fontes independentes de variação. Este é um caso comum quando as variáveis são altamente correlacionadas mas esta correlação não é perfeita.

3.2.1.4.1. POSSÍVEIS SINTOMAS E ALGUMAS MEDIDAS DE MULTICOLINEARIDADE:

- Pequenas mudanças nos dados causam grandes mudanças nas estimativas dos parâmetros;

- Coeficientes podem ter desvio padrão elevado e baixo nível de significância embora sejam conjuntamente significantes e o R^2 da regressão seja elevado;
- Coeficientes apresentam o sinal errado ou uma magnitude implausível.

A seguir são sugeridas algumas medidas de Multicolinearidade (GREENE, 1997):

1. Seja a matriz \mathbf{X} de dimensão K composta de uma coluna de 1's e as demais colunas representando $K-1$ variáveis medidas em desvios com relação à suas médias amostrais. Então, o primeiro elemento diagonal da matriz $(\mathbf{X}'\mathbf{X})^{-1}$ é igual a $1/(S_{11}(1-R_1^2))$, onde R_1^2 é o coeficiente de determinação da regressão de \mathbf{x}_1 em uma constante e todas as demais variáveis independentes. Conseqüentemente, a variância de b_1 é dada por:

$$\text{Var}(b_1) = \frac{\sigma^2}{(1-R_1^2)S_{11}} \quad (38)$$

Então, se uma variável altamente correlacionada com \mathbf{x}_1 for adicionada à Regressão, S_{11} não sofrerá mudança radical enquanto que o R_1^2 aumentará, assim como $\text{Var}(b_1)$. Porém, a regra de considerar as correlações simples entre as variáveis pode não dar uma indicação adequada do problema.

2. Uma outra sugestão é a medida proposta por LEAMER (1978)².

$$c_2(b_k) = \left\{ \frac{[\sum_i (x_{ik} - \bar{x}_k)^2]^{-1}}{[(\mathbf{X}'\mathbf{X})_{kk}^{-1}]} \right\}^{1/2} \quad (39)$$

² Lemaer, E. *Specification Searches: Ad Hoc Inferences with Nonexperimental Data*. New York: John Wiley and Sons, 1978.

que equivale à raiz quadrada da razão das verdadeiras variâncias de b_k estimadas sem e com as outras variáveis. Se \mathbf{x}_k é não correlacionado com as outras variáveis, o valor em (39) é 1. Caso contrário, este valor seria equivalente a $(1 - R_1^2)^{1/2}$, que sugere novamente a utilização dos coeficientes de determinação.

3. Uma terceira alternativa é comparar o coeficiente de determinação da regressão (R^2) com os R_i^2 's. Se o valor do R^2 for menor que um dos R_i^2 , então existe o problema da multicolinearidade nos dados.
4. Uma quarta sugestão é dada por BELSLEY et al. (1980)³, onde a grandeza do número de condição (raiz quadrada entre a máxima e a mínima raízes características) da matriz $\mathbf{X}'\mathbf{X}$ é analisada. Para este cálculo, ele sugere que a matriz $\mathbf{X}'\mathbf{X}$ deve ser normalizada, dividindo-se cada uma de suas colunas por $(\mathbf{x}'_k\mathbf{x}_k)^{1/2}$. Quanto maior o número de condição, maior a intercorrelação entre as variáveis. O autor sugere que valores acima de 20 são críticos.
5. BELSEY et al. (1980)² ainda sugerem uma outra medida de colinearidade. Para obtenção desta medida, parte-se do princípio de que a matriz de covariância do vetor obtido por Mínimos Quadrados pode ser escrita em termos de sua decomposição espectral:

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C}\mathbf{\Lambda}\mathbf{C}' = \sigma^2 \sum_{k=1}^K (1/\lambda_k) \mathbf{c}_k \mathbf{c}'_k \quad (40)$$

onde λ_k é a k -ésima raiz característica de $\mathbf{X}'\mathbf{X}$, estando os auto-valores ordenados em ordem decrescente, e \mathbf{c}_k é o vetor característico associado. Então, a variância do parâmetro b_k pode ser escrita da seguinte forma:

³ Belsley, D., E. Kuh and R. Welsch. *Regression diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley and Sons, 1980.

$$\text{Var}(b_k) = \sigma^2 \sum_{l=1}^k (1/\lambda_l) C_{kl}^2 \quad (41)$$

Cada variância pode, então, ser quebrada na proporção associada com cada raiz característica, através da seguinte expressão:

$$\pi_{kl} = \frac{C_{kl}^2 / \lambda_l}{(X'X)^{-1}_{kk}} \quad (42)$$

tal que $\sum_l \pi_{kl} = 1$. Os valores encontrados para os π 's são arranjados em uma tabela que revela dependência linear entre os regressores associados com uma raiz característica particular. Assim, raízes características pequenas sugerem altos valores para π_{kl} , tal que deve-se prestar atenção para grandes valores nas colunas da tabela, especialmente os que excedem 0,5. Teoria adicional sobre a obtenção dos auto-valores e seus respectivos auto-vetores pode ser encontrada em livros de Álgebra Linear como, por exemplo, LIPSCHULTZ (1968) e ANTON (1991).

3.2.1.4.2. POSSÍVEIS SOLUÇÕES PARA O PROBLEMA DA MULTICOLINEARIDADE

Duas soluções têm sido propostas com o objetivo de lidar com o mais óbvio problema da multicolinearidade: o alto desvio-padrão dos parâmetros. A primeira é a utilização do estimador de regressão *Ridge*, que é dado por:

$$\mathbf{b}_r = [\mathbf{X}'\mathbf{X} + r\mathbf{D}]^{-1}\mathbf{X}'\mathbf{y} \quad (43)$$

onde \mathbf{D} é uma matriz diagonal contendo os valores diagonais de $(\mathbf{X}'\mathbf{X})$ e r é escolhido arbitrariamente. Uma sugestão é iniciar com um valor pequeno de r como, por exemplo, 0,01 e ir aumentando-o gradativamente até que os coeficientes se estabilizem. Porém, o estimador *Ridge*, além de ser viciado, é de difícil inferência.

A segunda solução para o problema é o uso de *componentes principais*, cujo objetivo é o de extrair da matriz \mathbf{X} um pequeno número de variáveis que contribuem para a maior parte da variação em \mathbf{X} . Porém, o estimador obtido por este critério também é viciado e sua utilização pode apresentar problemas.

Apesar dos problemas já mencionados, este trabalho fará uso de Regressão *Ridge*, devido às maiores facilidades computacionais deste métodos.

3.2.1.5. HETEROCEDASTICIDADE

O problema de heterocedasticidade ocorre quando a variância do distúrbio não é constante. Quando isto ocorre, pode-se mostrar que o estimador do vetor \mathbf{b} obtido por Mínimos Quadrados é ineficiente e, portanto, outro tipo de estimador deve ser utilizado (GREENE, 1997). Além da simples análise visual, alguns testes podem ser elaborados com o objetivo de verificar heterocedasticidade, cujo objetivo é testar uma hipótese do tipo $H_0: \sigma_i^2 = \sigma$ para todo i .

O primeiro teste é o *Teste Geral de White*. Para obter a estatística de teste, basta calcular o valor de nR^2 na regressão de e_i^2 com todas as variáveis únicas da matriz \mathbf{X} . Esta estatística é assintoticamente distribuída com uma Distribuição Qui-Quadrado com $P-1$ graus de liberdade, onde P é o número de regressores, incluindo o termo constante. Este teste é extremamente geral e, embora possa revelar a existência de heterocedasticidade, não consegue identificar outros erros de especificação.

Outro teste sugerido é o *Teste de Goldfeld-Quandt*. Para este teste, assume-se que o grupo de dados possa ser dividido em dois grupos de maneira que, sob a hipótese de homocedasticidade, as variâncias dos distúrbios seriam iguais para os dois grupos. Para executar tal teste, CASSIDY (1981) sugere que as observações devem ser ordenadas pela variável suspeita de heterocedasticidade, que pode ser indicada através da análise dos gráficos de resíduos. A partir daí, o conjunto de dados deve ser dividido em três subconjuntos: o primeiro e o último com m observações e o segundo com $(n - 2m)$ observações. A estatística de teste é, então dada por:

$$F = \frac{SSE_1}{SSE_2} \quad (44)$$

onde SSE_1 e SSE_2 são as Somas dos Quadrados dos Resíduos obtidas das regressões do primeiro e do último conjuntos de observações, respectivamente. A estatística F apresenta distribuição F com $(m-K)$ e $(m-K)$ graus de liberdade, onde K é o número de parâmetros estimados.

3.2.1.6. OS EFEITOS DE INTERAÇÃO

Segundo GUNST & MASON (1980), os termos de interação em um modelo de regressão são produtos de duas ou mais variáveis preditoras. Tais termos são usuais quando se acredita que o efeito de uma variável preditora na resposta depende dos valores de outra(s) variável(is) preditora(s).

Uma forma de analisar este tipo de efeito é sugerida por GREENE (1997). Seja o modelo dado por:

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1 x_2 \quad (45)$$

Derivando- se y com relação a x_1 e x_2 , obter-se-á as seguintes relações:

$$\begin{aligned} \frac{\partial E(y)}{\partial x_1} &= \beta_2 + \beta_4 x_2 & e \\ \frac{\partial E(y)}{\partial x_2} &= \beta_3 + \beta_4 x_1 \end{aligned} \quad (46)$$

As relações mostradas em (46) implicam no fato de que, se β_4 é positivo, o efeito marginal de x_1 é aumentando quando x_2 aumenta. O mesmo ocorre quando se avalia o efeito marginal de x_2 . Para a construção de testes de hipóteses

e/ou intervalos de confiança para estes efeitos marginais, a variância é calculada conforme mostrado na expressão (47) para o efeito marginal de x_1 . Neste caso, um valor para x_2 deve ser escolhido. Uma escolha natural seria o valor amostral médio, embora em algumas circunstâncias, seja melhor utilizar o valor máximo obtido.

$$\text{Var}\left(\frac{\partial \hat{E}(y)}{\partial x_1}\right) = \text{Var}(\hat{\beta}_2) + x_2^2 \text{Var}(\hat{\beta}_4) + 2x_2 \text{Cov}(\hat{\beta}_2, \hat{\beta}_4) \quad (47)$$

3.3. OS SISTEMAS DE INFORMAÇÃO GEOGRÁFICA (SIG)

São várias as definições para os Sistemas de Informação Geográfica. De certa forma, o que muda de uma para outra são apenas as palavras, pois todas elas conduzem ao mesmo tipo de raciocínio. Somente para exemplificar, algumas destas definições são listadas a seguir.

Segundo CÂMARA et al. (1996), os SIGs podem ser definidos da seguinte maneira:

“Sistemas de Informação Geográfica – SIGs – são sistemas automatizados usados para armazenar, analisar e manipular dados geográficos, ou seja, dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação e indispensável para analisá-la”.

Segundo BUZAI & DURAN (1997), existem diferentes definições, que foram sendo dadas ao longo do desenvolvimento deste tipo de sistema. A primeira é sugerida por Dacey⁴:

⁴ Dacey, M. F. *Linguistics Aspects of maps and Geographic Information*. In: *Ontario Geography*, n. 5, Ontario, 1970.

“Qualquer coisa que funcione como um mapa, ao comunicar geograficamente a informação solicitada pelos usuários do sistema”.

Porém, os autores consideram tal definição ultrapassada, pois, desde o aparecimento da geoinformática, é possível armazenar, processar e reportar dados espaciais através da tecnologia computacional. Neste caso, ainda segundo BUZAI & DURAN (1997), torna-se mais adequada a definição de Marble⁵:

“É um sistema baseado em computação para o manejo de dados espaciais”.

Segundo HUXHOLD (1991):

“Há tantas definições quantas são as disciplinas envolvidas na utilização de Sistemas de Informação Geográfica, quais sejam: geografia, planejamento urbano, engenharia, processamento de dados, arquitetura, ciência ambiental e outras. Além disso, os SIG’s consistem de uma classe de ferramentas que profissionais nestas disciplinas usam para melhorar a maneira como trabalham. Assim como telefones, calculadoras e processadores de palavras são ferramentas para tornar o trabalho mais fácil, rápido e significativo, um SIG produz ferramentas para o profissional melhorar sua eficiência e eficácia no trabalho com informação em mapas e atributos de dados não gráficos”.

Diante do que se expôs, as definições dos SIG’s refletem a multiplicidade de usos e visões possíveis desta tecnologia e conduzem a um caráter interdisciplinar desta. Sendo assim, é possível indicar duas importantes

⁵ Marble, D. **Geographic Information Systems: An overview.** In: Peuquet, D. & Mable, D. (Eds.) *Introductory Readings in Geographic Information Systems.* Taylor & Francis. London. 1990.

características dos SIG's: a primeira é que estes sistemas possibilitam a integração, numa única base de dados, de informações geográficas provenientes de fontes diversas (dados cartográficos, dados de censo e cadastro urbano e rural, imagens de satélite e modelos numéricos de terreno) e a segunda é que os SIG's oferecem mecanismos para recuperar, manipular e atualizar estes dados, através de algoritmos específicos para estes propósitos.

3.3.1. HISTÓRICO

As primeiras tentativas de automatizar o processamento de dados georeferenciados ocorreram na Grã-Bretanha e nos USA, visando diminuir os custos de produção e atualização de mapas. No primeiro caso, o sistema foi desenvolvido para pesquisas em botânica, enquanto que no segundo o objetivo principal era a engenharia de tráfego (CÂMARA et al., 1996).

Somente nos anos 60 surgiram os primeiros SIG's propriamente ditos. Seu desenvolvimento teve origem no Canadá através de um plano estratégico governamental de longo prazo para criar um inventário automatizado de recursos naturais, com vistas ao planejamento destes recursos e ao uso do solo.

Durante os anos 70 desenvolveram-se fundamentos matemáticos voltados para a cartografia. Data desta década o surgimento da topologia aplicada, que permitiu análises espaciais entre elementos cartográficos.

Porém, os SIG's, que somente eram utilizados por grandes organizações até a década de 80, passaram a se difundir, o que ocorreu devido à popularização e barateamento das estações de trabalho, dos computadores pessoais e dos bancos de dados.

Este mesmo progresso tecnológico permitiu aperfeiçoar os mecanismos de aquisição de dados georeferenciados, o que aumentou a complexidade de coleta, armazenamento, manipulação e visualização dos dados, em função do seu volume, variedade e heterogeneidade.

CÂMARA⁶ considera três gerações de Sistemas de Informação Geográfica.

“A primeira, baseada em CAD cartográfico, caracteriza-se por sistemas herdeiros da tradição de Cartografia, com suporte de banco de dados limitado e cujo paradigma típico de trabalho é o mapa. Esta classe de sistemas é utilizada principalmente em projetos isolados, sem a preocupação de gerar arquivos digitais de dados. Esta geração também pode ser caracterizada como sistemas orientados a projeto.

A segunda geração de SIG's, baseada em banco de dados geográficos, chegou ao mercado no início da década de 90 e caracteriza-se por ser concebida para uso em ambientes cliente-servidor, acoplado a gerenciadores de bancos de dados relacionais e com pacotes adicionais para processamento de imagens. Desenvolvida em ambientes multiplataforma com interfaces em janelas, esta geração também pode ser vista como sistemas para suporte a instituições.

Para o final da década de 90, é previsto o aparecimento de uma terceira geração de SIG's baseada em bibliotecas digitais geográficas ou centros de dados geográficos, caracterizada pelo gerenciamento de grandes bases de dados geográficos, com acesso através de redes locais e remotas, públicas ou privadas. Para esta terceira geração, o crescimento de bancos de dados geográficos e a necessidade de seu compartilhamento com outras instituições requer o recurso de tecnologias como bancos de dados distribuídos e federativos. Estes Sistemas deverão seguir os requisitos de interoperabilidade, de maneira a permitir o acesso de informações espaciais por SIG's distintos. A terceira geração do SIG pode ainda ser vista como o desenvolvimento de sistemas orientados à troca de informações entre uma instituição e os demais componentes da sociedade”.

⁶ Camara G, *Modelos, Linguagens e Arquiteturas para Bancos de Dados Geográficos*. Tese de Doutorado, INPE, 1995.

3.3.2. TRATAMENTO DA INFORMAÇÃO

Conforme mencionado nas definições de SIG apresentadas anteriormente, o grande objetivo desta ferramenta, de forma bastante resumida, é associar um banco de dados alfanuméricos a um banco de dados gráficos. Porém, para que esta associação ocorra com sucesso, deve obedecer certos pré-requisitos com relação ao tratamento prévio dos respectivos dados de entrada.

3.3.2.1. TRATAMENTO DE DADOS ALFANUMÉRICOS

O tratamento de dados alfanuméricos pode ser realizado através do uso de editores de texto (que armazenam informações em ASCII), administradores de bancos de dados (que organizam os dados em forma matricial, onde as linhas representam os registros e as colunas representam as características de cada registro) e planilhas de cálculo. Uma outra alternativa são os programas de análise estatística, onde o tipo de tratamento de dados se baseia em dois grandes grupos: análise descritiva das distribuições e medição por agrupamentos (inclui diferentes tipos de análises numéricas e suas conseqüentes representações gráficas) e o da associação de fenômenos (dominado por diferentes formas de análises de associação através do uso de coeficientes de correlação – por exemplo, análise de *clusters*).

Outra opção é o uso de dados provenientes de sensores localizados em satélites artificiais, através do uso de GPS (sistema militar desenvolvido pelo Departamento de Defesa dos Estados Unidos que permite obter, mediante utilização de sinais emitidos por satélites artificiais e de receptores na terra, as coordenadas geográficas de qualquer ponto sobre a superfície terrestre), cujas medições podem, em certas circunstâncias, fornecer precisão inferior a um metro.

3.3.2.2. TRATAMENTO DOS DADOS GRÁFICOS

O tratamento de dados gráficos pode se dar através de Projeto Assistido por Computador (CAD – *Computer Aided Design*) ou Cartografia Assistida por Computador (CAC). A utilização deste tipo de ferramenta se realiza mediante a utilização de programas que servem para a digitalização de entidades gráficas vetoriais, basicamente pontos, arcos e polígonos que, combinados, formam um mapa denominado “vetorial”. Este tipo de programa possui grande capacidade e precisão no manejo com dados em duas ou três dimensões. Em geoinformática se aplica principalmente de maneira bidimensional para realizar cartografia assistida por computador (CAC), incorporando ao formato digital os mapas analógicos em papel mediante digitalização em mesa.

Uma segunda alternativa é o Processamento Digital de Imagens (PDI), que é realizado através da utilização de *software* utilizado em geoinformática especificamente para o tratamento de imagens digitais provenientes do uso de *scanners* ou satélites espaciais. O tipo de estrutura de imagens provenientes dos *scanners* e dos sensores incorporados aos satélites se denomina *raster* (ou matricial) e será discutido em maiores detalhes no próximo item.

A Modelagem Numérica de Terreno consiste de um procedimento computacional pelo qual se representa o espaço geográfico em três dimensões.

O tratamento combinado dos arquivos correspondentes aos dois tipos de bases de dados e sua georeferenciação pode ser executado através da tecnologia dos SIG's, cuja estrutura pode ser representada através do diagrama mostrado na Figura 3.7.

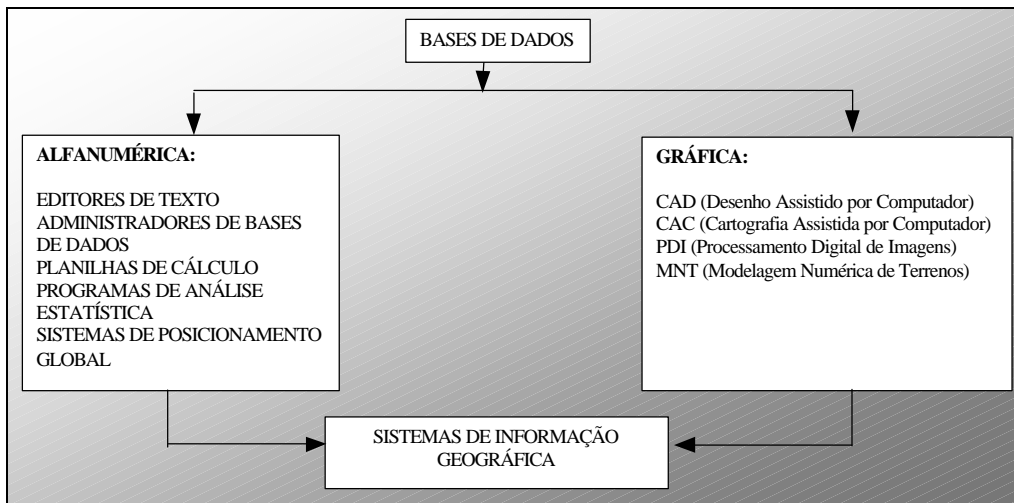


Figura 3.7 – Diagrama esquemático de um SIG.

3.3.3. ESTRUTURAS DE REPRESENTAÇÃO DO ESPAÇO GEOGRÁFICO

A forma como o *software* armazena e representa os dados na tela tem um importante papel no tipo de análise a ser conduzida. Dentre as estruturas de representação do espaço geográfico pelo SIG, duas têm se destacado: o modelo matricial e o modelo vetorial.

No modelo matricial, a representação gráfica é formada por uma matriz quadriculada, onde cada célula representa uma unidade mínima de resolução espacial: o *pixel*. A informação sobre o espaço geográfico que cada *pixel* contém é numérica e o número contido no *pixel* constitui uma das características descritivas em que se apresenta a composição cartográfica através de um grafismo ou de uma cor. A informação contida provém da associação de um número digital a cada *pixel* como unidade mínima de representação. Este valor é numérico e sua escala vai de zero (branco total) a 255 (preto total) em sua máxima possibilidade gráfica.

Desta forma, o tratamento matricial de cada mapa oferece vários tipos de representação para uma mesma área de estudo. A análise das informações se efetua por superposição de várias camadas temáticas. Um exemplo de modelo matricial ou *raster* é apresentado na Figura 3.8.

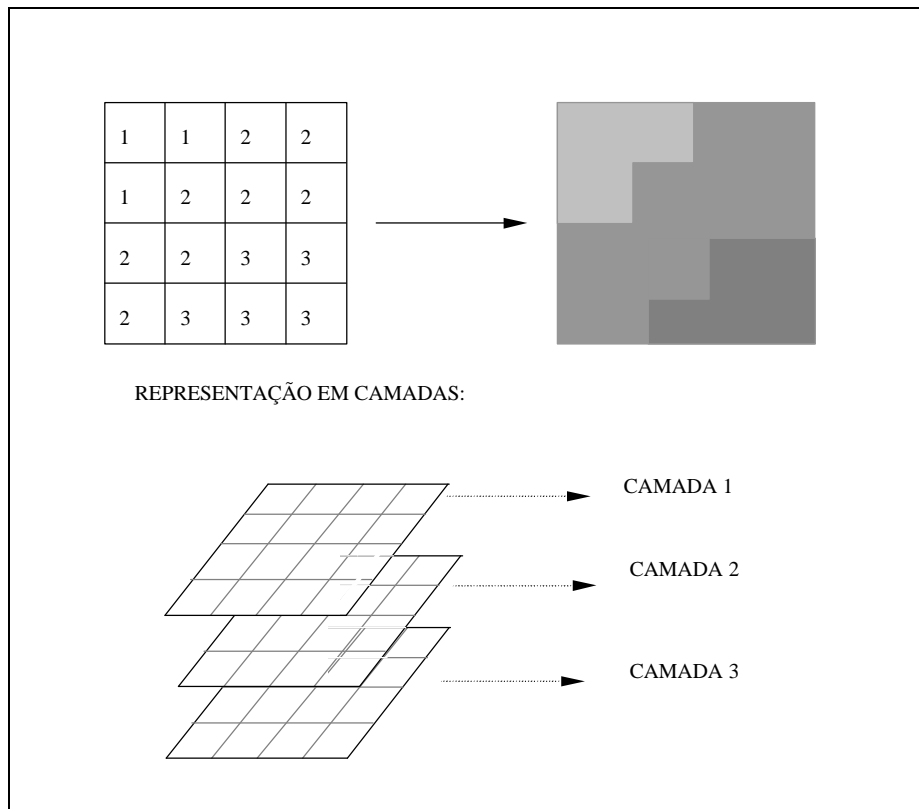


Figura 3.8 – Modelo matricial (ou *raster*).

O modelo *quadtree* é uma variante do modelo *raster*. Neste modelo, tem-se que em uma única composição cartográfica pode-se contar com diferentes tipos de *pixels*. Em localizações onde a homogeneidade espacial se mostra extensa, os *pixels* são maiores, enquanto que em locais onde as variações de categorias são muito pronunciadas, se usa células menores. O *quadtree* é uma estrutura de dados baseada na decomposição regular da imagem em quadrantes e subquadrantes. Esta estrutura divide hierarquicamente o espaço em determinados níveis de *quadtrees*. No nível 1, o espaço é dividido em quatro quadrantes. Se não há homogeneidade, cada um dos quadrantes do nível 1 é dividido em mais quatro quadrantes e assim por diante (TEIXEIRA et al., 1992).

O modelo vetorial opera representando o espaço geográfico de maneira contínua e se baseia nas entidades características de todo desenho cartográfico: o ponto, a linha e a área. As bases de dados utilizadas por este modelo são do tipo relacional, onde os campos contém a informação espacial e os registros

representam a totalidade das unidades espaciais consideradas. Uma base de dados geográfica difere da base convencional, uma vez que contém a localização espacial de cada registro e uma tabela topológica guarda suas relações. De maneira geral, em uma base de dados geográfica pode-se associar a informação alfanumérica à gráfica. Um modelo de uma base de dados vetorial é mostrado na Figura 3.9.

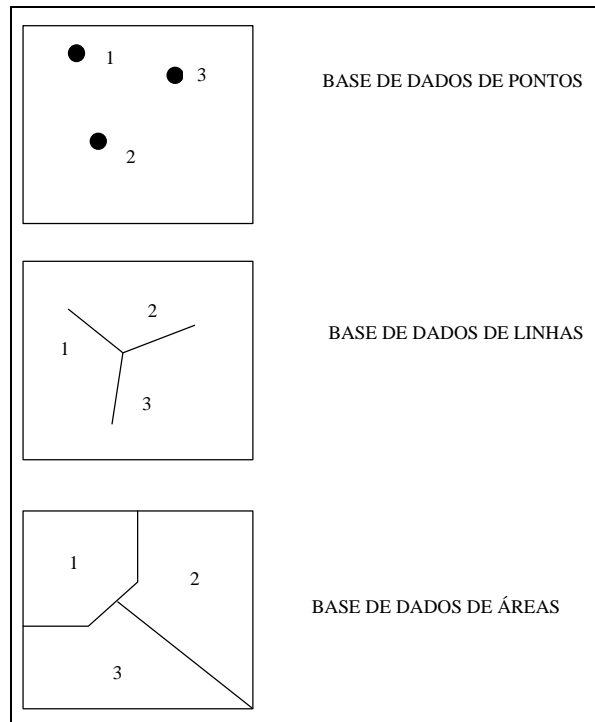


Figura 3.9 – Modelo vetorial.

3.3.4. DIFERENÇAS ENTRE OS SIG's E O CAD

Segundo SILVA et al. (1997), a principal diferença entre um SIG e um CAD reside no fato de que o primeiro é capaz de identificar as relações entre objetos, ou seja, a topologia. Um exemplo de como a topologia em sistemas deste tipo funciona, pode ser quanto à interpretação do desenho. Enquanto que um CAD interpreta um quadrado como a junção de quatro linhas, o SIG pode interpretar o mesmo quadrado como uma área, cujas características estão armazenadas em um banco de dados associado, o que torna possível realizar análises com as características espaciais dos objetos.

3.3.5. SITUAÇÃO ATUAL DO USO DE SIG NA AVALIAÇÃO IMOBILIÁRIA

Os SIG's vêm sendo utilizados nas mais diversas áreas de pesquisa, com resultados bastante satisfatórios. Dentre estas áreas, pode-se mencionar estudos recentes de avaliação imobiliária, que têm utilizado esta ferramenta como banco de dados para análises e como um fator de visualização espacial de características, valores reais e valores estimados.

Entre estes estudos, pode-se mencionar o trabalho de HADDAD (1995 e 1997), que consta de um estudo sobre o mercado imobiliário para a cidade de São Paulo feito a partir de dados de propriedades à venda. Em seu trabalho foi feito um estudo temporal-espacial da valorização destas propriedades, assim como dos pólos de valorização da cidade.

Em outro trabalho, levado a cabo por SILVA & WAERDEN (1997) para fins de taxaço, se estudou, através do uso de SIG, a aplicaço de dois modelos designados a calcular o custo adicional para levar redes de infra-estrutura e de transporte urbano a bairros localizados na periferia e separados da cidade por grandes vazios urbanos. Ainda com vistas à taxaço de propriedades, pode-se citar o relato de ULUACHA (1997), desenvolvido para a Nigéria.

BRONDINO & SILVA (1998a e 1998b) também utilizaram SIG em aplicaço de seus modelos de avaliaço para duas cidades brasileiras: São Carlos e Araçariguama, respectivamente.

Dentro da literatura brasileira, pode-se citar ainda outros casos de utilizaço de SIG's em trabalhos de avaliaço de imóveis, tais como DUARTE & GABBAY (1995), YUAÇA (1996), RAIA JR et al. (1996a e 1996b) e SILVA (1998). A utilizaço da ferramenta para este fim se intensifica devido ao fato de que em um SIG, "as propriedades de gráficos interativos permitem ao usuário gerar saídas gráficas rapidamente e resumir ou relacionar dados estatísticos num contexto geográfico. Esta técnica gráfica fornece um método efetivo de avaliaço dos resultados de análise, particularmente com grandes conjuntos de dados" (HSIAO &

STERLING, 1992). Esta propriedade do SIG pode ser ainda incrementada com o uso de um conjunto de técnicas específicas denominado estatística espacial, como se pode ver no próximo tópico.

3.3.6. ESTUDOS PRELIMINARES SOBRE ESTATÍSTICA ESPACIAL

“A estatística espacial é definida como uma coleção de técnicas para análise geográfica onde o resultado da análise depende do arranjo espacial dos eventos. Evento geográfico é entendido como uma coleção de pontos, linhas ou objetos, localizados no espaço, ligados com uma classe de atributos...” (HAINING, 1995).

Baseado na coleção sistemática de informação quantitativa, os objetivos da estatística espacial são:

- 1) Descrição cuidadosa e precisa de eventos no espaço geográfico (incluindo a descrição de padrões);
- 2) Exploração sistemática do padrão dos eventos e a associação entre eventos no espaço com o objetivo de ganhar um melhor entendimento dos processos que podem ser responsáveis pela distribuição dos eventos observada, e
- 3) Melhorar a habilidade de prever e controlar eventos que ocorrem no espaço geográfico.

O interesse por ferramentas de estatística espacial acopladas aos Sistemas de Informação Geográfica (SIG) tem crescido significativamente (LONGLEY & BATTY, 1996). Alguns SIG's já têm ferramentas deste tipo acopladas ao programa, como é o caso do *software* ARC INFO. Alguns pesquisadores, por sua vez, fazem ligações entre pacotes estatísticos e SIG's. Outros, ainda, desenvolvem *macros* específicas para este tipo de análise dentro do próprio SIG.

Os resultados obtidos até o presente momento sugerem que os grupos de métodos para os quais a ligação com o SIG vêm produzindo melhores resultados seriam aqueles das estatísticas descritivas simples e os relacionados com a estrutura de análise de covariância. Outras áreas da estatística tais como métodos de alisamento, funções-K e regressão espacial podem produzir benefícios equivalentes. Porém, até o presente momento, um progresso moderado tem sido em direção à visualização geográfica. Algum progresso também tem sido observado com relação a métodos estatísticos descritivos, o que conduz à impressão de que muita coisa ainda pode ser feita com relação ao “casamento” estatística espacial-SIG. Neste sentido, esforços podem ser destinados não ao desenvolvimento de ferramentas internas ao SIG mas sim na melhora da interface entre programas estatísticos e SIG’s.

Segundo LEVINE (1996), há algumas estatísticas espaciais que podem descrever a distribuição espacial dos dados. As variáveis que podem ser descritas espacialmente podem tanto ocorrer em localizações pontuais únicas (acidentes, construções, pessoas) ou estar agregadas (zonas de tráfego, limites da cidade etc.). As estatísticas que podem descrever tanto pontos quanto áreas pertencem a três categorias gerais:

1. Medidas de distribuição espacial: descrevem o centro, a dispersão, direção e forma da distribuição de uma variável;
2. Medidas de auto-correlação espacial: descrevem a relação entre as diferentes localizações para uma variável simples, indicando o grau de concentração ou dispersão, por exemplo, análise de *clusters*;
3. Medidas de associação espacial entre duas ou mais variáveis: descrevem a correlação ou associação entre variáveis distribuídas no espaço, por exemplo, correlacionar lojas de bebidas com lugares onde ocorrem acidentes.

O mesmo LEVINE (1996) recomenda alguns programas de computador já disponíveis para executar este tipo de análise de acordo com o tipo de problema a ser estudado.

Desta forma, utilizando estes conceitos, cada vez mais, pesquisadores têm feito uso desta técnica, que no entanto é ainda bastante incipiente. Um exemplo da utilização de estatística espacial em avaliação imobiliária vem do trabalho de BULLEN (1997). Em seu trabalho, modelos de regressão foram usados para estimar valores de propriedades no Reino Unido. A autora apresenta uma introdução à modelagem em múltiplos níveis (*Multilevel*), onde modelos diferentes de avaliação são estimados para diversas regiões. Os resíduos são, então, adicionados a um Sistema de Informação Geográfica, onde as regiões com preços abaixo e acima da média podem ser visualmente identificadas.

Ainda nesta direção, o trabalho de PATKAR (1997) sugere combinações possíveis de ferramentas matemáticas e SIG e o trabalho de WONG & WU (1995) alerta para o perigo da agregação, sugerindo alguns métodos para calcular os erros e distorções provocados por este procedimento.

Além dos métodos estatísticos convencionais para análise espacial, um estudo desenvolvido por WYATT (1997) faz uso de Redes Neurais (Kohonen) para fazer análise de agrupamentos.

Como pode ser observado por esta breve pesquisa bibliográfica, a estatística espacial se apresenta como uma ferramenta poderosa de análise em vários campos de estudo (RIPLEY, 1981). Porém, sua união com os SIG's está ainda em fase bastante inicial e se apresenta como uma ferramenta potencial de análise.

4. METODOLOGIA

Neste capítulo serão explicados os procedimentos adotados para o desenvolvimento deste trabalho. Na seqüência, haverá uma explanação sobre as cidades que foram objeto deste estudo e do processo de coleta de dados empregado para cada uma delas.

O método empregado para o desenvolvimento desta pesquisa foi desenvolvido com base nos principais objetivos deste trabalho que, de forma resumida, são os seguintes:

- Identificar variáveis importantes para a formação do valor da propriedade
- Utilizar Redes Neurais para estimar valores das propriedades
- Avaliar a influência da acessibilidade na valorização da propriedade urbana

Estabelecidos os objetivos, houve a necessidade de se identificar as fontes de dados necessárias. Para tal fim, as prefeituras de duas cidades foram contatadas: Araçariguama e São Carlos. A primeira, colocou à disposição um levantamento de características da maioria das propriedades locais, feito em 1995. Quanto à segunda, além dos dados do Censo de 1991, fornecidos pelo IBGE, a prefeitura colocou à disposição alguns dados constantes no setor de cadastro.

Escolhidas as cidades, o próximo passo foi decidir quais variáveis fariam parte do modelo de avaliação, com base na bibliografia pesquisada. Para a

cidade de Araçariгуama, as variáveis escolhidas ficaram limitadas aos principais aspectos constantes na ficha de cadastro. Para São Carlos, foi elaborada uma ficha com diversas características das propriedades a serem pesquisadas.

O próximo passo foi o de coleta da amostra. Para Araçariгуama, a amostra constou, praticamente, da totalidade de terrenos vazios situados em seis bairros. Para São Carlos, a amostra foi dimensionada com base no valor por m².

Os preços dos terrenos em Araçariгуama também foram fornecidos por um cadastro da prefeitura. Para São Carlos, características e preços de terrenos à venda foram levantados junto às imobiliárias.

Os terrenos amostrados foram, então, digitalizados nos mapas de ambas as cidades através de recursos do *software* TransCAD⁷, para que as características espaciais, como a acessibilidade, pudessem ser levantadas.

Construídos os conjuntos de dados, estes foram divididos, de forma aleatória, em três grupos distintos, destinados para cada uma das três fases do modelo de Redes Neurais, ou seja, treinamento (50% dos dados), validação e teste (25% dos dados em cada). Esta divisão foi feita três vezes para cada uma das cidades, gerando três grupos distintos de dados. Como, em geral, a definição do número de camadas escondidas e de neurônios em cada uma delas camadas se dá por experimentação, redes com várias configurações foram testadas, utilizando como ferramenta o *software* *Neural Planner*, versão 4.2. Cada configuração de rede foi testada para cada um dos três grupos de dados e o erro relativo total, cuja fórmula é dada pela equação (48), foi calculado em cada caso. Desta forma, gerou-se três valores de erro para cada topologia. O próximo passo constou do cálculo da média destes erros. Assim, o valor mínimo para esta média conduziu à melhor topologia, ou seja, à configuração de rede que forneceria estimativas mais próximas ao valor real observado.

$$\sum \text{erros} = \sum \left| \frac{o - e}{o} \right| \quad (48)$$

⁷ TransCAD é um programa criado e comercializado pela Caliper Corporation, 1172 Beacon Street, Newton, MA – 02161, USA.

onde o representa o valor observado e e , o valor estimado pelo modelo.

Além da média dos valores dos erros fornecidos pela equação (48), os valores do R^2 , obtido pelo mesmo critério utilizado na regressão, também foi utilizado para verificar a adequabilidade do modelo escolhido.

Para verificar a influência das variáveis no valor dos lotes, após escolhida a melhor configuração de rede, NATH et al. (1997) sugerem que a relevância de cada variável no modelo pode ser analisada segundo o método proposto por Garson⁸. Este método particiona os pesos sinápticos entre a camada intermediária e a de saída em componentes associados a cada nó de entrada. Assim, o peso resultante associado a cada entrada é um reflexo de sua importância.

Tal método é descrito aqui de forma resumida. Inicialmente, assume-se que w_{ij} ($i=1,\dots,k$; $j=1,\dots,p$) seja o peso que conecta o neurônio de entrada i com o neurônio j na camada intermediária. Da mesma forma, assume-se que w_{rs} ($r=1,\dots,p$) seja o peso que conecta o neurônio r da camada intermediária com o neurônio de saída. Então, o peso w_{rs} , independentemente de seu sinal pode ser incorporado aos pesos w_{ij} , usando a expressão (49), a seguir:

$$w_{ij}^* = \left\{ w_{ij} \mid / S_j \right\} (|W_{rs}|) \quad (49)$$

onde $S_j = \sum_{i=1}^k |w_{ij}|$.

A seguir, para cada nó de entrada, os pesos ajustados w_{ij}^* são somados para cada um dos nós da camada intermediária, ou seja, para todo i calcula-se $\sum_{j=1}^p w_{ij}^*$.

Desta forma, o percentual dado pela equação (50) serve como uma medida de saliência da variável representada pelo nó de entrada.

⁸ Garson, D. G. *Interpreting neural-network connection weights*. AI Expert, 1991, April, 47-51.

$$\text{relevância da variável } i \text{ (\%)} = \frac{\sum_{j=1}^p w_{ij}^*}{\sum_{i=1}^k \sum_{j=1}^p w_{ij}^*} \quad (50)$$

Uma vez identificadas as variáveis relevantes, uma nova rede é treinada considerando-se apenas estas variáveis como entrada. O percentual de representação de uma variável que a torne irrelevante depende do autor e do assunto. Para este estudo em particular, foram retiradas as variáveis que apresentavam percentuais de representação iguais ou menores a 5%.

O próximo passo constou da comparação espacial dos resultados, objetivando verificar se os erros obtidos pelos métodos empregados possuíam algum padrão de distribuição espacial. Esta comparação de resultados partiu da idéia de verificar em quais bairros os modelos não apresentavam boas estimativas e, se possível, corrigir a distorção existente, através da adição de algum elemento novo ao modelo. A primeira alternativa para este tipo de estudo seria a utilização de um programa específico para análise espacial, conforme o sugerido por LEVINE (1996). Porém, como não havia disponibilidade imediata deste tipo de *software*, resolveu-se utilizar um método misto, onde técnicas estatísticas de agrupamento foram utilizadas e o resultado obtido foi adicionado ao SIG.

Para a construção dos mapas temáticos em questão, os erros relativos obtidos para cada um dos três conjuntos, utilizando o Modelo de Regressão como padrão, foram agrupados pelo método dos k-vizinhos mais próximos. Para cada conjunto, os erros foram divididos em quatro *clusters*, cujos limites também foram utilizados para construir os intervalos dos erros obtidos pelas ANN. A seguir, verificou-se o percentual de pontos que se enquadrava em cada *cluster* para ambos os métodos.

4.1. A CIDADE DE ARAÇARIGUAMA

Araçariguama é uma cidade localizada a aproximadamente 40 km de São Paulo e foi, durante muitos anos, distrito do município de São Roque e somente

há pouco tempo conseguiu sua emancipação. Como característica principal do lugar, pode-se citar que ela é uma cidade do tipo “dormitório”, ou seja, boa parte de sua população economicamente ativa trabalha nas cidades vizinhas, principalmente, São Paulo, São Roque e Sorocaba.

Atualmente, o número de habitantes no lugar está em torno de 6000 e o número de propriedades, aproximadamente 1500, entre terrenos e construções. A cidade está dividida em nove bairros, a saber: Centro, Vila Real, Vale da Benção, Chácara D’ora, Terra Baixa, Vila Alumínio, Cintra Gordinho, Estância Imperial e Vila Nova.

4.1.1 COLETA DE DADOS

Por ocasião do levantamento de dados para este projeto, a Prefeitura local havia iniciado os trabalhos de cadastramento das propriedades e possuía registros apenas daquelas localizadas em seis dos bairros citados: Centro, Vila Real, Vale da Benção, Chácara D’ora, Estância Imperial e Vila Nova. Dentre as características constantes nos arquivos originais deste cadastro, procurou-se retirar as mais relevantes para construir o modelo de avaliação, com base na bibliografia pesquisada.

As variáveis retiradas do cadastro, acompanhadas das codificações empregadas na construção do banco de dados, são dadas a seguir:

- **FORMA:** forma do terreno - variável binária, assumindo valor 1 (regular) ou 0 (irregular);
- **TESTADA:** testada principal do terreno (m);
- **ÁREA:** área do terreno (m²);
- **MURO:** existência de muro - variável binária, assumindo valor 1 (presença) ou 0 (ausência);
- **CALÇADA:** existência de calçada - variável binária, assumindo valor 1 (presença) ou 0 (ausência);

- **SITUAÇÃO:** situação dentro da quadra - variável binária, assumindo valor 1 (esquina) ou 0 (outros);

A sétima variável empregada, menor caminho percorrido até o centro através da malha viária em quilômetros (MC), foi obtida através de uma rotina do *software* TransCAD.

Os preços dos imóveis foram obtidos, também de cadastro da Prefeitura, por um ex-aluno de Engenharia Civil desta Escola, oriundo da cidade em questão, como parte de um trabalho de graduação.

Diante do que se expôs, pode-se dizer que não houve, para este projeto, no caso de Araçariguama, um trabalho prévio de dimensionamento da amostra, nem um critério de amostragem específico. Porém, a amostra é constituída de 157 lotes, ou seja, mais que 10% do total de endereços (contando-se construções e terrenos vazios) da cidade. Para os seis bairros citados, a amostra é constituída praticamente da totalidade da população de imóveis não construídos.

Um mapa da cidade, onde podem ser vistos os terrenos constantes da amostra, é apresentado na Figura 4.1.

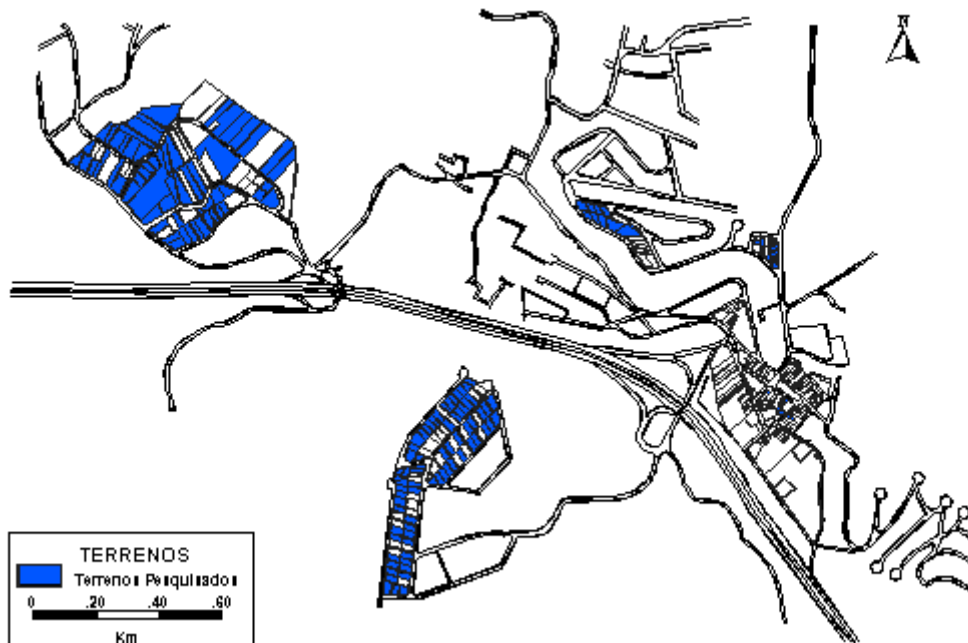


Figura 4.1 - Distribuição espacial dos lotes pesquisados na cidade de Araçariguama.

4.2 A CIDADE DE SÃO CARLOS

São Carlos situa-se na região centro-oeste do estado de São Paulo, distante 230 km da capital. Possui cerca de 160.000 habitantes e conta hoje com um importante centro industrial, quando comparada às demais cidades do interior com mesmo porte. Possui empresas como Johann Faber (utensílios para desenho e escrita), Tecumseh (compressores), Eletrolux Prosdócimo (eletrodomésticos), Latina (eletrodomésticos) e Volkswagen (automotiva). Possui ainda várias empresas de pequeno e médio porte, que atuam na área de máquinas agrícolas, componentes para as empresas de porte maior e informática. Conta ainda com o Parque de Alta Tecnologia, estrutura montada para fornecer apoio a novos empreendimentos ligados a tecnologias de ponta.

Possui também duas Universidades públicas, um *campus* da Universidade de São Paulo e o *campus* da Universidade Federal de São Carlos, que somados à EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) tornam a cidade um importante centro de pesquisas.

4.2.1 COLETA DE DADOS

Como se pode notar pelos modelos citados anteriormente, cada modelo de avaliação pode fazer uso dos mais diversos tipos de variáveis. Cada lugar apresenta características próprias de valorização e deve ser tratado individualmente. Um modelo que produz bons resultados para uma cidade, pode não fornecer boas estimativas para outra, mesmo estando as duas muito próximas fisicamente.

Para a coleta de dados em São Carlos, recorreu-se inicialmente ao Setor de Cadastro da Prefeitura Municipal, onde os fechamentos de quadra, que são uma espécie de planta de cada quarteirão, foram obtidos. Nestes fechamentos constam todos os imóveis pertencentes à quadra, assim como os respectivos endereços. Tais dados, totalizando mais de 2800 quadras, foram utilizados para a execução do plano de amostragem e para a localização dos lotes amostrados no mapa da cidade.

Para a escolha das variáveis que fariam parte deste estudo, foram consultados alguns especialistas na área imobiliária em São Carlos, a fim de verificar quais são os fatores que mais pesam para o comprador na aquisição de um imóvel. De posse destas informações, de algumas variáveis utilizadas em estudos similares e das variáveis sugeridas pela Norma NBR 5676, elaborou-se uma ficha para a execução da pesquisa de campo. As variáveis constantes nesta ficha, acompanhadas de suas codificações para entrada no banco de dados, são dadas a seguir:

- **ÁREA:** área do terreno (m²);
- **TESTADA:** testada principal do terreno (m);
- **TOPOGRAFIA:** declividade do terreno – variável binária, assumindo valor 1 (plano) ou 0 (outros);
- **FORMA:** forma do terreno - variável binária, assumindo valor 1 (regular) ou 0 (irregular);
- **SITUAÇÃO:** situação dentro da quadra - variável binária, assumindo valor 1 (esquina) ou 0 (outros);
- **EDIF:** existência de edificações na quadra – variável binária assumindo valor 1 (sim) ou 0 (não);
- **ÁGUA:** existência de rede de água tratada – variável binária assumindo valor 1 (presente) ou 0 (ausente);
- **ESGOTO:** existência de rede de esgoto – variável binária assumindo valor 1 (presente) ou 0 (ausente);
- **LUZ:** existência de rede de energia elétrica - variável binária assumindo valor 1 (presente) ou 0 (ausente);
- **ASFALTO:** existência de pavimentação– variável binária assumindo valor 1 (presente) ou 0 (ausente);
- **MURO:** existência de muro - variável binária assumindo valor 1 (presente) ou 0 (ausente)

- CALÇADA: existência de calçada - variável binária assumindo valor 1 (presente) ou 0 (ausente)

As características físicas obtidas foram lançadas num Sistema de Informações Geográficas. A medida de acessibilidade utilizada, ou seja, a distância ao centro através da malha viária dada em quilômetros (MC), também foi obtida a partir dos recursos do SIG utilizado, neste caso, o *software* TransCAD.

A princípio, duas eram as possibilidades para a aquisição de dados: uma junto às imobiliárias locais e outra junto ao Cartório de Registro de Imóveis da cidade. A segunda opção foi logo descartada porque, conforme informações de corretores, em geral, as escrituras na cidade são registradas utilizando-se o valor venal do imóvel fornecido pela Prefeitura, que é bastante baixo comparado ao valor de mercado, tornando esta informação muito pouco confiável. Optou-se, então, por procurar os dados nas imobiliárias.

Segundo alguns profissionais do setor imobiliário consultados em São Carlos, há atualmente na cidade uma grande oferta de terrenos, devido ao último plano de estabilização monetária do governo federal (Plano Real). Quem adquiriu um terreno antes do Plano Real para vendê-lo mais tarde a um preço mais elevado teve uma surpresa: ninguém quer comprar. Estas pessoas querem agora vender os imóveis a um preço que seja, no mínimo, igual ao preço que pagaram. Acontece que este preço ainda está acima do valor que os compradores aceitam pagar hoje, fazendo com que haja hoje uma grande oferta de terrenos, mas com preços muito altos. Portanto, os preços dos terrenos a serem utilizados serão aqueles a princípio almejados pelos proprietários destes e que muitas vezes estão acima do valor de mercado e acabam sendo vendidos com descontos.

Além disso, infelizmente, a maioria das imobiliárias não dispõe de todos os dados desejados sobre as características dos lotes. Quase todas as imobiliárias organizam os dados em fichas e cadernos, pouquíssimas organizam em computadores pessoais e, em algumas, somente o corretor tem conhecimentos sobre os dados dos terrenos, o que dificulta muito a coleta. Outro fator que atrapalha a amostragem é a ausência de exclusividade de venda do terreno pela imobiliária. O

dono do terreno pode colocá-lo à venda em quantas imobiliárias ele quiser, por isso quase não há placas em terrenos que estão à venda, impossibilitando uma coleta de dados *in loco* antes de realizar a pesquisa em imobiliárias. No caso de haver dados do mesmo terreno em mais que uma imobiliária, assumiu-se como seu preço a média entre todos os valores pedidos por ele.

4.2.1.1 DIMENSIONAMENTO DA AMOSTRA

Para a coleta de dados, a cidade foi dividida em três grupos diferentes de terrenos. Esta escolha deu-se pelo fato da cidade possuir propriedades com os mais variados valores por metro quadrado e características. Optou-se, então, por dividir a cidade em bairros de três categorias, a saber: A, B e C, sendo os do tipo A, os bairros que apresentam o m² mais caro. Com esta divisão, além de conseguir bons resultados com amostras menores, pode-se obter modelos mais precisos em termos de avaliação. Um modelo que contivesse todos os tipos de propriedades, implicaria numa grande variabilidade de valores e poderia subestimar ou superestimar o valor de algumas. Para esse fim, foram consultados especialistas na área de mercado imobiliário e, a partir dos dados fornecidos por estes e do mapeamento de renda obtido pelos dados do Censo de 1991, a divisão em classes foi elaborada como consta nas Tabelas 4.1, 4.2 e 4.3. Como bairros do tipo A, foram considerados aqueles cujos terrenos apresentassem valores maiores que R\$ 50,00 por m². Como bairros do tipo B, classificou-se aqueles onde o valor por metro quadrado está situado na faixa de R\$ 36,00 a R\$ 50,00 e do tipo C, os de valor por m² abaixo de R\$ 36,00.

Tabela 4.1 - Classificação dos bairros da cidade de São Carlos pela classe de preço A.

Centro	Chácara São João
Cidade Jardim	Jardim Bethania
Jardim Macarengo	Jardim Ricetti
Jardim Santa Helena	Jardim Santa Paula
Parque Arnaldo Schimid	Parque Estância Suíça
Parque Nova Estância Suíça	Parque Sabará
Parque Santa Mônica	Vila Elisabeth
Vila Monteiro	Vila Nery
Jardim Nova Santa Paula	Jardim São Carlos
Loteamento Albertini	Loteamento Centreville
Loteamento Taba Yacy	Vila Prado

Tabela 4.2 - Classificação dos bairros da cidade de São Carlos pela classe de preço B.

Chácara São Caetano	Jardim Bandeirantes
Jardim Brasil	Jardim Centenário
Jardim Gilbertoni	Jardim Hikari
Jardim Luthfalla	Jardim Nossa Senhora Aparecida
Jardim Nossa Senhora de Fátima	Jardim Nova São Carlos
Jardim Paulista	Jardim Paulistano
Morada dos Deuses	Núcleo Residencial Silvio Villari
Parque Industrial Miguel Abdelnur	Parque Paraíso
Parque Primavera	Parque São José
Portal do Sol	Recreio dos Bandeiraantes
Solar do Engenheiros	Tijuco Preto
Vila Arnaldo	Vila Bela Vista
Vila Boa Vista	Vila Brazilia
Vila Costa do Sol	Vila Marina
Vila Pelicano	Vila Pureza
Vila Rancho Velho	Vila Santo Antônio
Vila São José	Vila Vista Alegre
Jardim Alvorada	Jardim Cardinalli
Parque Santa Mônica II	Vila Deriggi
Planalto Paraíso	Vila Faria
Jardim São João Batista	Jardim São Paulo
Vila Alpes	Vila Marcelino
Vila Marigo	Vila Morumbi
Vila Redenção	

Tabela 4.3 - Classificação dos bairros da cidade de São Carlos pela classe de preço C.

Chácara Bataglia	Chácara Monjolinho
Chácara Paraíso	Chácaras de Recreio Mont Carlo
Cidade Aracy	Jardim Beatriz
Jardim Botafogo	Jardim Cruzeiro do Sul
Jardim das Rosas	Jardim Gonzaga
Jardim Guanabara	Jardim Jockey Clube
Jardim Maracanã	Jardim Maria Alice
Jardim Medeiros	Jardim Pacaembu
Jardim Paraíso	Jardim Real
Jardim Santa Maria	Jardim Santa Tereza
Jardim São Gabriel	Jardim Tangará
Loteamento D'Aquino	Núcleo Residencial Castello Branco
Parque Delta	Parque Delta II
Parque Industrial	Parque Sisi
Recreio São Judas Tadeu	Residencial Itamaraty
Santa Felícia	São Carlos II
São Carlos III	São Carlos IV
Vila Carmen	Vila Celina
Vila Conceição	Vila Isabel
Vila Jacobucci	Vila Monte Carlo
Vila Santa Madre Cabrini	Vila Santa Maria
Vila Marques	

Estimação do tamanho da amostra:

O primeiro passo foi definir o tamanho da amostra que iria ser retirada de cada grupo. Uma escolha natural seria a de que o tamanho da amostra deveria ser proporcional à população total do grupo a que ela pertencia.

Pelo critério de amostragem sugerido por COCHRAN (1953), pode-se dimensionar o tamanho da amostra e, conseqüentemente o de cada grupo, a partir de uma variância “desejada”, V , que pode ser obtida através de estudos prévios ou situações parecidas. Esta variância pode, ainda, ser obtida através de um coeficiente de variação, estimado como sendo igual a um percentual do valor da média. Como não havia, pelo menos nos últimos anos, algum trabalho em avaliação de terrenos através do qual uma estimativa para esta variância pudesse ser obtida, foi feita uma estimativa grosseira do valor da média dos terrenos para toda a população e atribuiu-se um desvio padrão, aceitável, que fosse igual a 20% do valor desta média.

Em termos quantitativos, para uma média global igual a R\$43,63, foi assumido um desvio padrão como sendo 20% deste valor, ou seja, igual a 0,20 que multiplicado por 43,63, resultou em 8,73. Foi através deste desvio-padrão “desejado”, que o dimensionamento da amostra foi obtido. O valor estimado desta média global foi obtido a partir de informações fornecidas por algumas imobiliárias, onde faixas de valores por m² para terrenos situados em alguns dos bairros das categorias A, B e C foram obtidas.

De posse do valor de V, o tamanho total da amostra, n, foi obtido através da equação (51):

$$n = \frac{n_0}{1 + \frac{1}{V} \sum N_h s_h^2} \quad (51)$$

onde: $n_0 = \frac{(\sum N_h s_h)^2}{V}$;

s_h^2 : variância amostral do grupo h;

N_h : tamanho da população do grupo h.

Sendo n_h o tamanho da amostra que vai ser retirada do grupo h, este número vai ser dado por:

$$n_h = \frac{n W_h s_h}{\sum W_h s_h} \quad (52)$$

onde: $W_h = \frac{N_h}{N}$

Através dos registros armazenados nos fechamentos de quadra no Setor de Cadastro da cidade fez-se a contagem de quantos terrenos sem construção existem em São Carlos, ou seja, o número total de lotes, N, chegando-se a um total de 15052. A partir daí, estas propriedades foram classificadas de acordo com as classes citadas nas Tabelas 4.1, 4.2 e 4.3, chegando-se às características de cada estrato, conforme mostrado na Tabela 4.4. Os valores das variâncias utilizadas também foram calculados com base em informações de corretores, que forneceram, *a priori*, a faixa de preços por m², em vários bairros das classes A, B e C da cidade.

Tabela 4.4 - Características das categorias.

GRUPO	TAMANHO DO GRUPO (N _h)	VARIÂNCIA DOS VALORES POR M ²	DESVIO PADRÃO DOS VALORES POR M ² (s _h)	W _h = N _h /N
A	1701	2116	46,00	0.11
B	3790	11,36	3,37	0.25
C	9561	7,10	7,10	0.64

A dimensão total da amostra foi obtida pela equação (51), utilizando-se os valores da Tabela 4.4, chegando-se a um valor de n igual a 2641. Os tamanhos das amostras para cada grupo foram obtidos pela equação (52) e são dados por 1301, 212 e 1128, para os estratos A, B e C, respectivamente.

Com o objetivo de verificar a adequabilidade do tamanho da amostra proposta por COCHRAN (1953), outra estimativa foi feita com base na fórmula para dimensionamento de n, proposta por LEVY & LEMESHOW (1991).

“Suponha que desejamos determinar o número de elementos necessários para ter 100 x (1 - α)% de certeza de obter, da amostra aleatória estratificada, uma média estimada, \bar{x}_{str} que difere da verdadeira média \bar{X} em não mais que 100 x ε%... A fórmula para o n requerido é como segue”:

$$n \approx \frac{\left(\frac{z_{1-\alpha/2}^2}{N^2} \right) \left(\sum_{h=1}^L \frac{N_h^2 \sigma_{hx}^2}{p_h \bar{X}^2} \right)}{e^2 + \left(\frac{z_{1-\alpha/2}^2}{N^2} \right) \left(\sum_{h=1}^L \frac{N_h \sigma_{hx}^2}{\bar{X}^2} \right)} \quad (53)$$

onde: N: tamanho populacional

N_h: tamanho do grupo h

z: valor tabelado da Distribuição Normal

\bar{X}^2 : média global

σ_{hx}^2 : variância do grupo h

$\pi_h = n_h/n$

Utilizando os mesmos dados do caso anterior, com $\varepsilon=20\%$, o valor de n obtido pela fórmula (53) vai ser igual a 1637, ou seja, menor que o obtido anteriormente. Os tamanhos das amostras a serem retiradas dos grupos A, B e C ficaram com as seguintes dimensões: 809, 131 e 697, respectivamente.

Calculados os tamanhos amostrais, iniciou-se a coleta de dados. Não foi possível, entretanto, obter o tamanho da amostra desejado a princípio pois toda a população de terrenos à venda na cidade de São Carlos, tanto os ofertados em imobiliárias quanto em jornais, esgotou-se, totalizando 480 lotes amostrados. Porém, isso não se torna um problema, pois, a maioria dos autores que publica estudos sobre avaliação utiliza apenas uma amostra dentre os imóveis que estão à venda. Somente para exemplificar, pode-se citar aqui os seguintes trabalhos: GONZÁLEZ (1995 e 1996), GONZÁLEZ & FORMOSO (1994), VERTELO (1996), TRIVELLONI et al. (1996) e BRANDÃO et al. (1995). O mapa para a cidade de São Carlos, com os terrenos amostrados é mostrado na Figura 4.2.

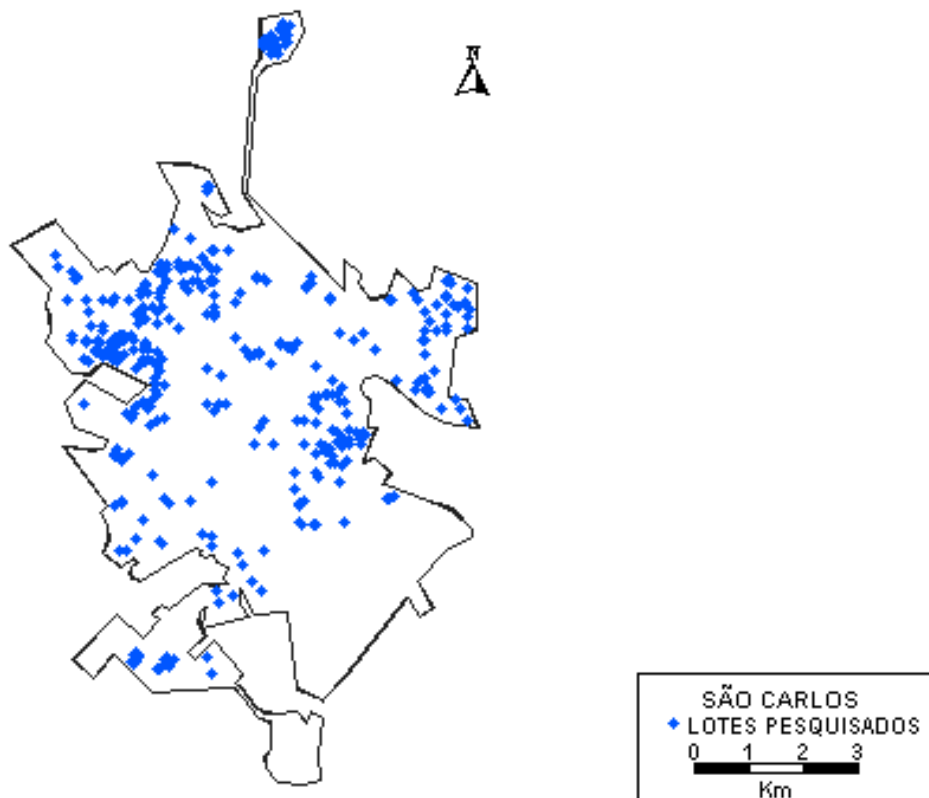


Figura 4.2- Distribuição espacial dos lotes pesquisados na cidade de São Carlos.

5. RESULTADOS

Este capítulo inicia com os resultados relativos ao primeiro estudo de caso deste trabalho: Araçariguama. Serão apresentados o modelo de regressão estimado, os resultados obtidos com a aplicação das Redes Neurais e uma comparação numérica e espacial do desempenho dos dois métodos. Em seguida, é mostrado o estudo de caso para a cidade de São Carlos, seguindo a mesma estrutura de apresentação.

5.1. ESTUDO DE CASO: ARAÇARIGUAMA

5.1.1. MODELO DE REGRESSÃO PARA A CIDADE DE ARAÇARIGUAMA

A escolha de um modelo adequado para a cidade de Araçariguama obedeceu a um processo de tentativa e erro, onde vários modelos foram testados, a fim de obter aquele que melhor se adequasse ao conjunto de dados.

Inicialmente ajustou-se um modelo do tipo linear aos dados, considerando-se o preço por m² como variável resposta. O valor 0,53 encontrado para o coeficiente R², mostrou que o modelo ajustado não era satisfatório. Tentou-se, então, adicionar fatores de interação entre área e testada, área e menor caminho e

testada e menor caminho, obtendo-se uma melhora significativa no coeficiente de determinação, que passou de 0,53 para 0,72. Quando a interação destes três fatores foi acrescentada a este modelo, no entanto, este coeficiente sofreu um aumento desprezível, em torno de 0,01. Além disso, tal modelo não obedeceu às suposições de Normalidade e variância constante do distúrbio aleatório.

A alternativa encontrada foi fazer novas tentativas, empregando-se transformações aos dados numéricos. Para tal, foram feitos gráficos de preço/m² versus área, preço/m² versus testada e preço/m² versus menor caminho. Posteriormente, foram ajustadas curvas aos pontos de cada um destes gráficos, cujos melhores resultados foram obtidos através das seguintes relações:

Preço/m² = função (área^{-0.5}), cujo valor para o R² foi aproximadamente igual ao obtido para a relação Preço/m² = função (ln (área))

Preço/m² = função (testada^{-0.6}), cujo valor para o R² foi aproximadamente igual ao obtido para a relação Preço/m² = função (ln (testada))

Preço/m² = função (mc + mc²)

Devido à maior facilidade de análise, a relação do tipo logarítmica foi escolhida para introduzir no modelo as variáveis área e testada. Já para a variável mc, com o objetivo de facilitar a interpretação dos resultados, seguiu-se a sugestão de CASSIDY (1980), que recomenda que apenas um dos termos seja adicionado ao modelo, preferencialmente o termo quadrático. Porém, quando o modelo geral foi testado, este termo conduziu a um menor valor para o R², com relação ao modelo que só continha esta variável em sua forma pura.

Os resultados obtidos para as melhores tentativas são mostrados na Tabela 5.1, onde foram incluídas apenas as equações que obtiveram um R² maior que 0,30.

Tabela 5.1 - Melhores modelos de regressão ajustados, considerando o preço por m^2 para a cidade de Araçariguama.

	MODELO	R²
1	(Preço/ m^2) = função (área, testada, forma, muro, calçada, menor caminho, situação)	0,53
2	(Preço/ m^2) = função (área, testada, forma, muro, calçada, menor caminho, situação, área*testada, área*mc, testada*mc)	0,72
3	(Preço/ m^2) = função (área, testada, forma, muro, calçada, menor caminho, situação, área*testada, área*mc, testada*mc, área*testada*mc)	0,73
4	(Preço/ m^2) = função ((área) ^{-0.5} , testada, forma, muro, calçada, menor caminho, situação, área*testada, área*mc, testada*mc)	0,74
5	(Preço/ m^2) = função ((área) ^{-0.5} , (testada) ^{-0.6} , forma, muro, calçada, menor caminho, situação, área*testada, área*mc, testada*mc)	0,76
6	(Preço/ m^2) = função ((área) ^{-0.5} , (testada) ^{-0.6} , forma, muro, calçada, (mc) ² , situação, área*testada, área*mc, testada*mc)	0,72
7	(Preço/ m^2) = função (mc, forma, situação, muro, calçada, (área) ^{-0.5} , testada ^{-0.6} , (mc) ² , área*testada, área*mc, testada*mc)	0,81
8	(Preço/ m^2) = função (mc, forma, situação, muro, calçada, ln (área), ln (testada), área*testada, área*mc, testada*mc)	0,81

Escolhida a função a ser utilizada (número 8 na Tabela 5.1), uma análise da multicolinearidade envolvendo as variáveis independentes foi executada. O número de condição obtido para a matriz normalizada $X'X$ foi igual a 46,07, indicando a possível existência de dependência linear entre duas ou mais variáveis. Para identificar as variáveis envolvidas em multicolinearidade, os valores do R_{ij}^2 π_{ij} foram calculados para cada par de variáveis. Os resultados obtidos são mostrados nas Tabelas 5.2 e 5.3, respectivamente.

Tabela 5.2 - Valores de π_{ij} obtidos para as colunas da matriz $X'X$ para Araçariguama.

	MC	FORMA	SITUAÇÃO	MURO	LN (ÁREA)	CALÇADA	LN (TESTADA)
MC	0,21	0,37	0,00	0,01	0,01	0,34	0,06
FORMA	0,27	0,32	0,27	0,00	0,09	0,04	0,00
SITUAÇÃO	0,03	0,03	0,61	0,15	0,13	0,04	0,00
MURO	0,00	0,01	0,00	0,39	0,49	0,01	0,12
LN (ÁREA)	0,00	0,00	0,00	0,00	0,00	0,03	0,97
CALÇADA	0,00	0,02	0,02	0,20	0,19	0,53	0,02
LN (TESTADA)	0,00	0,01	0,14	0,35	0,48	0,01	0,01

Tabela 5.3 - Valores dos R_{ij}^2 obtidos para cada par de variáveis para a cidade de Araçariguama.

	MC	FORMA	SITUAÇÃO	MURO	CALCADA	LN (ÁREA)	LN (TESTADA)
MC	1,00						
FORMA	0,03	1,00					
SITUAÇÃO	0,00	0,08	1,00				
MURO	0,00	0,02	0,00	1,00			
CALCADA	0,01	0,00	0,00	0,08	1,00		
LN (ÁREA)	0,07	0,07	0,00	0,01	0,01	1,00	
LN (TESTADA)	0,01	0,04	0,01	0,02	0,00	0,56	1,00

Pode-se observar pela Tabela 5.2 que as variáveis ln (área) e ln (testada) apresentaram um valor para o π_{ij} maior que 0,5, o que poderia ser indicativo de uma possível fonte de multicolinearidade. A Tabela 5.3 revela que o valor do R_{ij}^2 para esta relação não é maior que aquele obtido para o R^2 geral. Porém, segundo GREENE (1997), a regra utilizada para comparar os R_{ij}^2 com o R^2 geral é apenas sugestiva e "...não dá indicação de como proceder". Desta forma, optou-se por utilizar os valores da Tabela 5.2 para a tomada de decisão. Para resolver este problema, duas eram as possíveis soluções. A primeira seria a execução de uma regressão do tipo *Ridge*, o que geraria estimadores viciados e um conseqüente comprometimento do resultado dos testes de hipóteses. A outra alternativa seria a retirada de uma destas duas variáveis do modelo, assim como dos fatores de interação envolvendo-as, o que poderia conduzir à perda de informação. Optou-se, então, por elaborar as duas tentativas, verificar os resultados e, a partir daí, tomar a decisão sobre a escolha do modelo.

Para a regressão *Ridge* iniciou-se com um valor de r igual a 0,01, que foi sendo acrescido em 0,01 até que parâmetros estáveis fossem obtidos. A estabilidade no valor dos parâmetros foi obtida com $r=0,04$. O modelo final apresentou um R^2 igual a 0,70 e os parâmetros estimados são apresentados na Tabela 5.4.

Tabela 5.4 - Modelo de Regressão *Ridge* obtido para Araçariguama utilizando o valor do m^2 como variável dependente.

	PARÂMETRO ESTIMADO	DESVIO PADRÃO	ESTATÍSTICA t	p-level
CONSTANTE	144,49	13,1578	10,98132	7,93E-21
MC	-12,5288	1,832527	-6,83692	2,05E-10
FORMA	-5,88981	1,566124	-3,76075	0,000245
SITUAÇÃO	2,74178	2,122308	1,291886	0,198438
MURO	-2,4022	2,983959	-0,80504	0,422107
LN (ÁREA)	-11,3232	2,155853	-5,25229	5,21E-07
AR*MC	0,002285	0,000449	5,084042	1,12E-06
CALÇADA	11,2199	9,237964	1,214543	0,226501
TEST*MC	0,149612	0,047678	3,137973	0,002058
LN (TEST)	-11,858	3,476079	-3,41133	0,000836
AREA*TES	-2,1E-05	1,08E-05	-1,95433	0,052573

A tomada de decisão sobre qual das duas variáveis retirar do modelo foi baseada no valor do R^2 . Na primeira regressão, cujo coeficiente de determinação obtido foi igual a 0,67, tanto a variável ln (área) quanto os fatores de interação envolvendo-a foram retirados. A segunda tentativa constou em retirar a variável ln (testada), assim como seus fatores de interação. Nesta nova situação, cujos parâmetros podem ser vistos na Tabela 5.5, o R^2 obtido foi igual a 0,72, daí sua escolha.

Tabela 5.5 - Modelo de Regressão obtido para Araçariçuama utilizando o valor do m^2 como variável dependente após a retirada da variável \ln (testada).

	PARÂMETRO ESTIMADO	DESVIO-PADRÃO	ESTATÍSTICA t	p-level
CONSTANTE	146,4912	13,62418	10,75229	2,56E-20
MC	-10,0447	1,873369	-5,36181	3,08E-07
FORMA	-6,49788	1,545863	-4,2034	4,51E-05
SITUAÇÃO	2,753884	2,055459	1,33979	0,182353
MURO	-5,13412	2,9238	-1,75597	0,081148
LN (ÁREA)	-16,7819	2,436357	-6,88809	1,48E-10
ÁREA * MC	0,002686	0,000387	6,938565	1,13E-10
CALÇADA	15,77541	9,091707	1,735143	0,084783

Uma comparação entre as Tabelas 5.4 e 5.5 mostra que os parâmetros obtidos para as variáveis muro, \ln (área) e calçada sofreram as maiores alterações relativas. Pode-se verificar também que tanto o parâmetro estimado para a variável muro quanto para as variáveis calçada e situação não são significativos e, portanto, as alterações observadas não são um problema maior. Por outro lado, o aumento no coeficiente da variável \ln (área) pode ter decorrido da retirada do \ln (testada) do modelo.

Além do valor dos parâmetros, uma comparação entre as somas dos Quadrados dos Resíduos (SSE) para os dois casos foi feita com o objetivo de verificar qual das duas situações gerou menor erro de estimação. Para a regressão *Ridge*, a SSE obtida foi igual a 11560,98 contra um valor igual a 10984,75 para a outra alternativa.

Diante dos resultados obtidos, optou-se pelo modelo mostrado na Tabela 5.5, uma vez que, além de não ter sido notada nenhuma mudança significativa com relação ao outro modelo, pode-se fazer inferências mais seguras com relação aos parâmetros estimados.

Como os testes com relação aos parâmetros só são eficientes se o distúrbio aleatório possuir Distribuição Normal e variância constante, estas suposições deveriam ser verificadas a princípio. Assim, com o objetivo de verificar a suposição de Distribuição Normal, foi elaborado um teste do tipo Qui-quadrado para verificar a distribuição do resíduo da regressão. A estatística de teste obtida foi

igual a 11,02, cujo valor foi comparado ao tabelado a um nível de significância de 5%, donde a suposição de distribuição Normal não foi rejeitada.

Para analisar a suposição de heterocedasticidade foi elaborado, a princípio, o *Teste Geral de White*. A estatística de teste obtida, com valor igual a $2,59 \times 10^{-14}$, foi comparada com o valor tabelado da Distribuição Qui-quadrado para um valor crítico igual a 95% e 7 graus de liberdade (14,07). Como o valor calculado é menor que o tabelado, a hipótese de homocedasticidade não foi rejeitada.

Outro teste elaborado foi o *Teste de Goldfeld-Quandt*. Para tal, a amostra foi ordenada de acordo com o valor da variável mc e dividida em três grupos, o primeiro e o último com 55 pontos e o intermediário com 37 pontos. A hipótese de que a variância no primeiro grupo seria maior que a do último foi rejeitada a um nível de significância igual a 5% (estatística de teste obtida foi igual a 0,337 - cujo valor foi comparado ao valor tabelado da Distribuição F para um valor crítico igual a 95% com $n_1=47$ e $n_2=47$ graus de liberdade). Portanto, por este teste, a hipótese de homocedasticidade também não foi rejeitada.

Obedecidas as suposições de homocedasticidade e Distribuição Normal, pôde-se, então, fazer inferências sobre os parâmetros do modelo. Os dados da Tabela 5.5 revelam que, além do termo constante, os parâmetros obtidos para as variáveis mc, forma, ln (área) e para a interação entre a área e a distância são significativos. Desta forma, pode-se notar que a forma irregular causa decréscimo no valor do m^2 . Por sua vez, a área e a distância ao centro não podem ser tratadas de forma individual. Assim, pode-se verificar que o efeito marginal (proporcional a $0,0027 \times \text{área}$) da variável distância ao centro sobre o preço do m^2 sofre um ligeiro aumento quando a área aumenta, ou seja, áreas grandes e localizadas longe da região central têm seu preço aumentado. Por sua vez, o efeito marginal da variável área sobre o valor estimado não é tão fácil de ser medido, uma vez que este vai ser proporcional a $[(1/\text{área}) + 0,0027 \times mc]$.

5.1.2. REDES NEURAIS PARA A CIDADE DE ARAÇARIGUAMA

Inicialmente, o conjunto de dados foi subdividido, de forma aleatória, em três subconjuntos: o primeiro para treinamento da rede, o segundo para validação e o terceiro para teste, num total de 77, 40 e 40 lotes, respectivamente. Esta subdivisão foi executada três vezes, gerando três conjuntos diferentes de dados.

O tipo de topologia empregada foi a MLP (*Multilayer Perceptron*) com uma camada escondida e o algoritmo de aprendizado escolhido foi *backpropagation*, com atualização dos pesos por padrão.

Inicialmente, os dados não sofreram nenhum tipo de pré-processamento e foram montadas diversas redes, variando a taxa de aprendizado, o *momentum* e o número de nós na camada intermediária. Como os resultados obtidos não se mostraram satisfatórios, optou-se por normalizar os dados entre 0 e 1 e, posteriormente, entre 0,1 e 0,9, variando-se também a taxa de aprendizado e o *momentum* para estes dados. Os resultados das melhores tentativas, com a inclusão de todas as variáveis, são mostrados na Tabela 5.6.

Para cada um dos três conjuntos de treinamento, foram calculados os somatórios dos erros de previsão de preços, cujos valores foram obtidos através da equação (35) e que também são mostrados na Tabela 5.6. A última coluna da tabela refere-se à média dos somatórios dos erros obtidos.

Tabela 5.6 - Resultados das melhores simulações, usando todas as variáveis para a cidade de Araçariçuama.

	MODELO	nº neurônios na camada intermediária	taxa aprendizado	<i>Momentum</i>	Média dos erros
1	Dados brutos	3	0,3	0,3	28,84
2	Dados brutos	7	0,5	0,8	28,79
3	Dados brutos	4	0,5	0,8	31,99
4	dados brutos	7	0,3	0,8	30,65
5	dados brutos	7	0,2	0,8	30,02
6	dados brutos	7	0,4	0,4	27,48
7	dados brutos	7	0,3	0,3	29,48
8	norm. 0 -1	7	0,4	0,3	29,29
9	norm. 0 -1	7	0,5	0,8	31,08
10	norm 0,1- 0,9	4	0,3	0,4	33,22
11	norm 0,1- 0,9	7	0,3	0,4	17,47
12	norm 0,1- 0,9	4	0,3	0,4	16,91
13	norm 0,1- 0,9	4 e 3 (2 cam. int.)	0,3	0,4	16,00
14	norm 0,1- 0,9	3 e 2 (2 cam. int.)	0,3	0,4	16,37

Pode-se verificar pelos valores acima que, com os dados normalizados entre 0,1 e 0,9, a rede apresentou erro menor. Observa-se, também, que os melhores valores para a taxa de aprendizado e o *momentum* ficam na escala 0,3 a 0,4.

Devido ao fato de que o método de saliência de variáveis proposto por Garson somente pode ser empregado em redes com apenas uma camada intermediária, a configuração número 12 foi escolhida para o seu emprego, embora as redes 13 e 14 tenham apresentado erro ligeiramente menor. A Tabela 5.7 apresenta a saliência calculada para cada variável em cada um dos três conjuntos distintos de dados.

Tabela 5.7 - Relevância de cada variável para a cidade de Araçariçuama.

CONJUNTO 1		CONJUNTO 2		CONJUNTO 3	
VARIÁVEL	RELEV.(%)	VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)
MC	34	MC	59	MC	41
ÁREA	18	ÁREA	10	ÁREA	18
TESTADA	18	SITUAÇÃO	9	TESTADA	13
CALÇADA	16	TESTADA	7	SITUAÇÃO	7,5
FORMA	5	FORMA	6	FORMA	6
MURO	5	MURO	5	MURO	6
SITUAÇÃO	5	CALÇADA	4	CALÇADA	6

Embora a cidade em questão seja de pequeno porte, os dados da Tabela 5.7 revelam que a variável menor caminho exerce a maior influência sobre o valor estimado para os três conjuntos de dados, seguida pela variável área. Ao contrário do modelo de regressão, a variável testada aparece em terceiro lugar em dois dos conjuntos. Pode-se notar também que as variáveis forma e muro exercem pouca influência sobre os valores estimados. Conforme proposto na Metodologia apresentada aqui, as variáveis cuja influência fosse menor que 5% deveriam ser retiradas do conjunto de dados e haveria, então, um novo treinamento da rede. Portanto, diante dos valores observados na Tabela 5.7, optou-se, então, por retirar somente a variável muro, embora esta tenha mostrado saliência superior a este valor para o último conjunto de dados. A mesma configuração de rede foi então treinada para esta nova situação e a média dos erros obtida baixou para 16,54.

5.1.3. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELOS DOIS MÉTODOS PARA A CIDADE DE ARAÇARIGUAMA

Os mesmos conjuntos de dados utilizados para a parte de teste das Redes Neurais foram utilizados para estimar os erros obtidos pelo modelo de regressão. Além destes erros, o R^2 também foi calculado para os dois modelos. Estes resultados são mostrados na Tabela 5.8.

Tabela 5.8 - Erros e R^2 obtidos para os dois modelos e os três conjuntos de dados para Araçariguama.

	REGRESSÃO		RN	
	ERRO	R^2	ERRO	R^2
conjunto 1	29,49	0,23	18,77	0,23
conjunto 2	18,99	0,55	12,88	0,45
conjunto 3	22,61	0,62	17,98	0,79
MÉDIA	23,70	-	16,54	-

Os resultados mostrados na Tabela 5.8 indicam que o erro relativo total obtido pelo modelo de regressão ficou em torno de 43% maior que aquele fornecido pelas Redes Neurais. Quanto ao valor do R^2 , não se pode concluir muita

coisa, uma vez que os valores obtidos oscilaram muito de um conjunto de dados para outro.

Ainda com o objetivo de comparar os erros relativos obtidos através da equação (48) para os dois métodos, foram feitos os gráficos mostrados nas Figuras 5.1, 5.2 e 5.3 mostrando estes valores para ambos os métodos e os conjuntos de dados 1, 2 e 3, respectivamente.

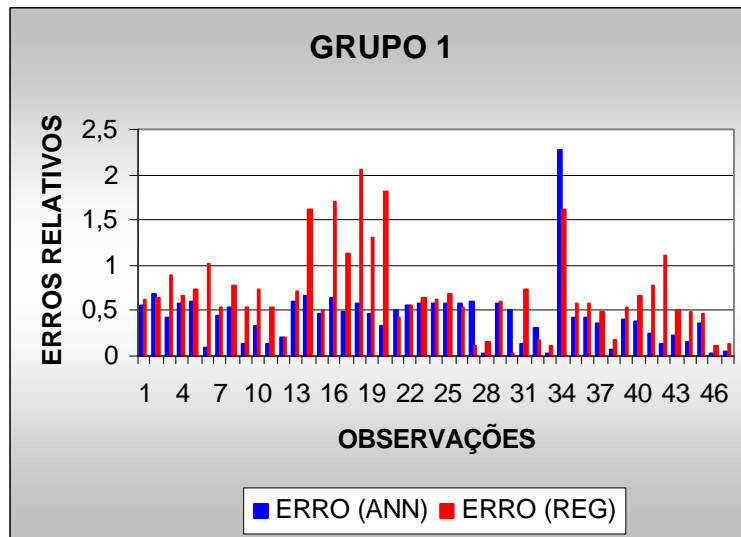


Figura 5.1 - Erros relativos obtidos para os dois métodos e o primeiro conjunto de dados para a cidade de Araçariguama.

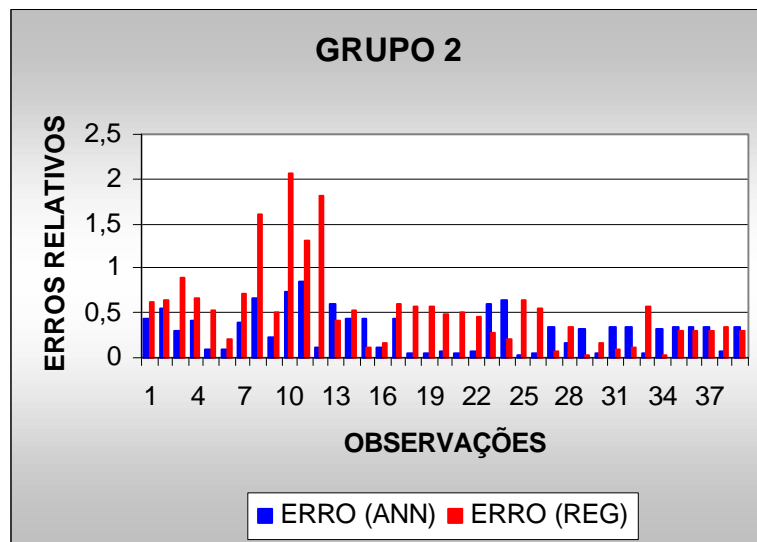


Figura 5.2 - Erros relativos obtidos para os dois métodos e o segundo conjunto de dados para a cidade de Araçariguama.

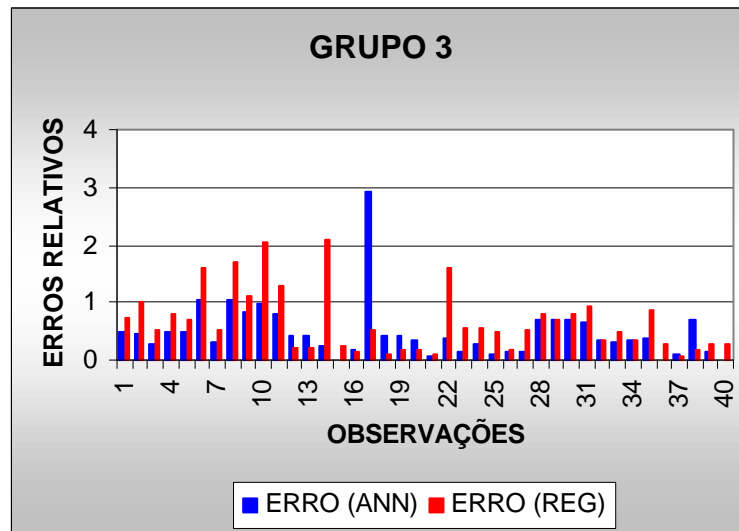


Figura 5.3 - Erros relativos obtidos para os dois métodos e o terceiro conjunto de dados para a cidade de Araçari.

Os gráficos mostrados nas Figuras 5.1, 5.2 e 5.3 indicam que, quando observados ponto a ponto, os resultados obtidos com as Redes Neurais apresentaram, em geral, erro relativo menor. Pode-se notar que para o primeiro e o segundo conjuntos, a maioria dos valores fornecidos por este método situou-se em torno de 0,5, podendo-se notar a existência de apenas um ponto discrepante para o primeiro conjunto. Para o terceiro grupo de dados, a situação é um pouco diferente, pois pode-se notar que os erros obtidos são um pouco maiores do que os observados anteriormente.

Num próximo passo, erros encontrados para o modelo de regressão para cada um dos três conjuntos foram agrupados em cinco *clusters* através da utilização do pelo método denominado *K-means clustering*. Os mesmos extremos foram, então, utilizados para agrupar os erros fornecidos pelas Redes Neurais. A proporção de pontos em cada categoria foi identificada (Tabela 5.9) e utilizada para a comparação dos resultados.

Tabela 5.9 - Proporção de erros relativos nos clusters encontrados com os dois modelos para os três conjuntos de dados para Araçariguama.

Conjunto/ <i>Cluster</i>	Limites de classe	REGRESSÃO	RN
Conjunto 1 - <i>Cluster</i> 1	$\leq 0,209$	19,0%	25,5%
Conjunto 1 - <i>Cluster</i> 2	0,209 a 0,602	32,5%	66,0%
Conjunto 1 - <i>Cluster</i> 3	0,603 a 0,890	30,0%	6,5%
Conjunto 1 - <i>Cluster</i> 4	0,891 a 1,302	8,5%	0,0%
Conjunto 1 - <i>Cluster</i> 5	$\geq 1,303$	10,5%	2,0%
Conjunto 2 - <i>Cluster</i> 1	$\leq 0,160$	21,5%	32,0%
Conjunto 2 - <i>Cluster</i> 2	0,161 a 0,348	24,0%	16,0%
Conjunto 2 - <i>Cluster</i> 3	0,349 a 0,602	27,0%	35,0%
Conjunto 2 - <i>Cluster</i> 4	0,603 a 0,890	16,0%	16,0%
Conjunto 2 - <i>Cluster</i> 5	$\geq 0,891$	11,0%	0,0%
Conjunto 3 - <i>Cluster</i> 1	$\leq 0,224$	25,0%	25,0%
Conjunto 3 - <i>Cluster</i> 2	0,225 a 0,359	15,0%	17,5%
Conjunto 3 - <i>Cluster</i> 3	0,360 a 0,719	27,5%	40,0%
Conjunto 3 - <i>Cluster</i> 4	0,720 a 1,302	20,0%	12,5%
Conjunto 3 - <i>Cluster</i> 5	$\geq 1,303$	12,5%	2,5%

Os resultados mostrados na Tabela 5.9 indicam que para os três conjuntos de dados, a concentração de pontos localizados nos intervalos que apresentam erros menores (*clusters* 1, 2 e 3) aumentou quando as RN foram utilizadas. Disso, pode-se concluir que, de maneira geral, o erro de estimação diminuiu quando esta técnica foi utilizada. Outra informação importante que pode ser retirada da mesma tabela é a diferença entre a concentração de pontos no *cluster* 5 para os dois métodos. Enquanto para a regressão este valor foi sempre superior a 10%, para as RN ele nunca superou os 2,5%.

Identificadas as proporções de pontos em cada *cluster*, partiu-se para a representação espacial dos resultados, através da construção de mapas temáticos. As Figuras 5.4, 5.5, 5.6, 5.7, 5.8 e 5.9 mostram como os erros de estimação estão distribuídos espacialmente para ambos os métodos e os três conjuntos de dados.

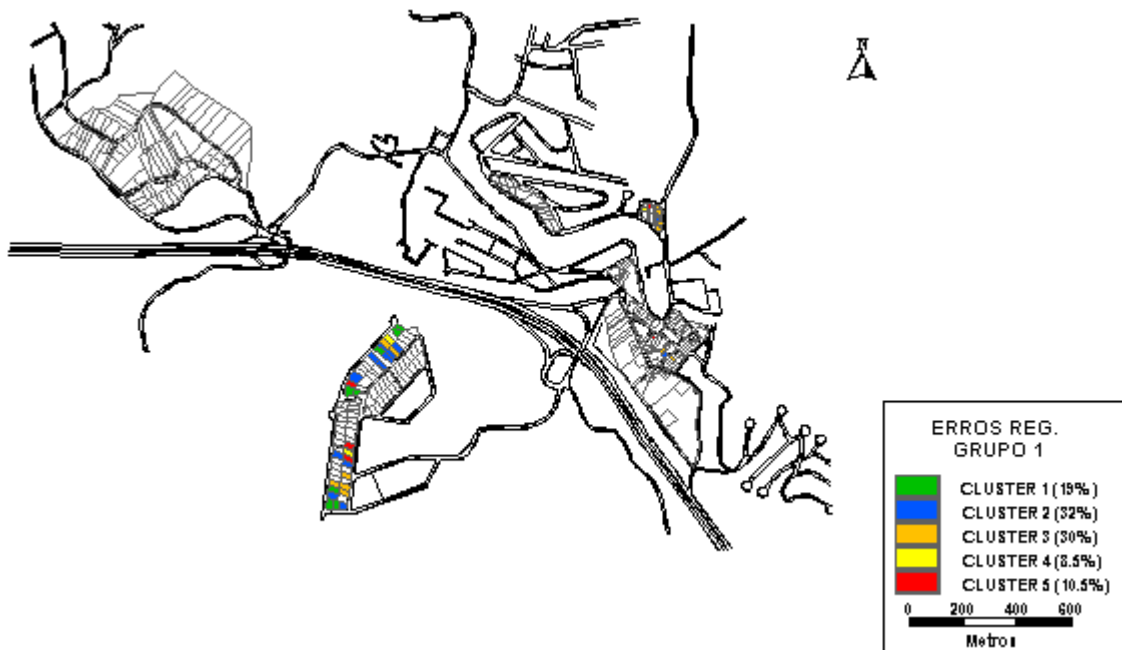


Figura 5.4 – Distribuição espacial dos erros para o modelo de regressão e o primeiro grupo de dados de Araçariçuama.

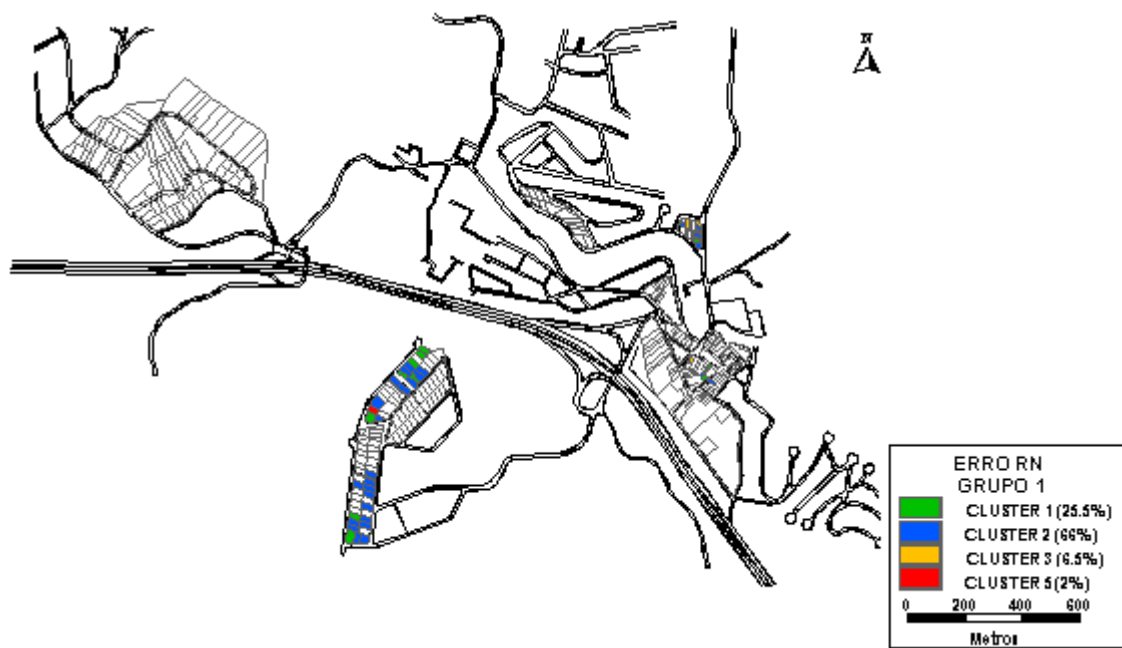


Figura 5.5 – Distribuição espacial dos erros para as RN e o primeiro grupo de dados de Araçariçuama.



Figura 5.6 – Distribuição espacial dos erros para o modelo de regressão e o segundo grupo de dados de Araçariguama.



Figura 5.7 – Distribuição espacial dos erros para as RN e o segundo grupo de dados de Araçariguama.



Figura 5.8 – Distribuição espacial dos erros para o modelo de regressão e o terceiro grupo de dados de Araçariguama.



Figura 5.9 – Distribuição espacial dos erros para as RN e o terceiro grupo de dados de Araçariguama.

A Figura 5.4 revela que a maioria dos pontos cujo erro de estimação é maior se concentra no bairro localizado abaixo da estrada. Neste mesmo bairro também se localiza o único ponto do *cluster 5* para as RN (Figura 5.5). Para os mapas mostrados nas Figuras 5.6 e 5.7, pode-se observar uma maior concentração

de valores grandes de erro no bairro localizado no lado superior esquerdo para ambos os modelos. Para o terceiro conjunto de dados e o modelo de regressão (Figura 5.8), no entanto, observa-se valores maiores de erros espalhados pelos dois bairros. Já para as RN, o único ponto pertencente ao *cluster* 5 também se localiza no bairro situado abaixo da estrada. Diante do que foi observado, pode-se dizer que a distribuição espacial dos erros não apresenta um padrão de localização específico para a Regressão. Por outro lado, observa-se que os maiores erros fornecidos pelas RN se localizam num bairro específico para dois dos conjuntos de dados (1 e 3).

5.2. ESTUDO DE CASO: SÃO CARLOS

5.2.1. MODELO DE REGRESSÃO PARA A CIDADE DE SÃO CARLOS

A metodologia para obtenção do Modelo de Regressão de São Carlos foi a mesma utilizada no Modelo de Regressão de Araçariguama. A única diferença está no maior número de variáveis qualitativas e maior número de registros.

Inicialmente, optou-se por trabalhar com o preço por m² como variável resposta. O primeiro modelo ajustado foi o linear, onde foram incluídas todas as variáveis independentes, mas sem considerar os fatores de interação. O valor de R² obtido foi 0,34.

Ao fazer uma transformação logarítmica na variável dependente, o R² elevou-se para 0,51. A partir daí, outras tentativas de transformações tanto na variável dependente quanto nas independentes também não conduziram a grandes melhoras (ver Tabela 5.10).

Diante destes valores, optou-se por fazer gráficos de (preço/m²) *versus* área, (preço/m²) *versus* testada e (preço/m²) *versus* menor caminho, a fim de verificar possíveis relações através de regressões simples. As melhores relações obtidas para as variáveis duas a duas são mostradas a seguir:

Preço/m² = função (área^{0,50})

Preço/m² = função (testada^{0,50}) e Preço/m² = função (e^{testada*0,02}) – que apresentaram aproximadamente o mesmo R².

Preço/m² = função (e^{-0,2*mc}) e Preço/m² = função (mc^{-0,2}) – que apresentaram aproximadamente o mesmo R².

A inclusão destas transformações ao modelo também não trouxe melhoras substanciais em termos de R², cujo valor passou para 0,43. Outras transformações nas variáveis independentes e também na dependente foram feitas, não conduzindo a nenhuma melhora significativa, em termos de coeficiente de determinação, como pode ser observado na Tabela 5.10 (na qual só são apresentados os modelos que forneceram R² maior que 0,30). A inclusão da categoria a que pertence o lote no modelo, contudo, gerou uma melhora substancial. Para a inclusão desta variável, adicionou-se ao modelo duas variáveis binárias: TERR_B (assumindo valor 1 para terrenos da categoria B e 0, caso contrário) e TERR_C (assumindo valor 1 para terrenos da categoria C e 0, caso contrário). Desta forma, terrenos pertencentes à categoria A possuíam valor (0,0) para as variáveis TERR_B e TERR_C. O novo modelo, que incluiu somente as variáveis em sua forma pura e sem transformação na variável dependente conduziu a um R² igual a 0,57. O modelo que incluiu uma transformação do tipo logarítmica na variável dependente gerou um coeficiente de determinação igual a 0,73 e, portanto, foi o escolhido para a análise.

De maneira similar à escolha do modelo para Araçariguama, uma análise da multicolinearidade envolvendo as variáveis independentes foi executada após a escolha do modelo. O número de condição obtido para a matriz normalizada X'X foi igual a 158,02, indicando a possível existência de dependência linear entre duas ou mais variáveis. Novamente, para identificar as variáveis envolvidas em multicolinearidade, os valores do π_{ij} e do R_{ij}² foram calculados para cada par de variáveis. Os resultados obtidos são mostrados nas Tabelas 5.11 e 5.12, respectivamente.

A Tabela 5.11 indica valores maiores que 0,5 para as seguintes relações: área e testada, testada e categoria, calçada e categoria, asfalto e testada, edificação e asfalto e água e luz. Pelos valores observados na Tabela 5.12 verifica-se que as relações área e testada e água e luz apresentam coeficientes de determinação maiores que o obtido para o modelo geral (0,73) e iguais a 0,87 e 0,80, respectivamente, indicando possíveis fontes de multicolinearidade. Novamente, duas eram as possíveis soluções: uma regressão do tipo *Ridge* ou a retirada de algumas variáveis e nova estimação do modelo. Da mesma forma que para Araçariguama, as duas tentativas foram analisadas.

Para a regressão do tipo *Ridge*, iniciou-se com um valor igual a 0,01 para r , que sendo foi acrescido em 0,01 até que os parâmetros atingissem a estabilidade. O modelo obtido, com $r=0,01$ e R^2 igual a 0,72 praticamente não difere daquele em que o tipo de modelagem convencional foi utilizado e é mostrado na Tabela 5.13.

Pelos mesmos motivos mencionados no estudo de caso para Araçariguama, optou-se por utilizar os valores dos π_{ij} para a retirada de variáveis envolvidas em multicolinearidade. Desta forma, optou-se por retirar a variável testada, uma vez que esta aparecia em três relações cujo valor para o π_{ij} era maior que 0,5, mantendo-se, então as variáveis categoria, área e asfalto. Outra relação que poderia ser considerada como possível fonte de multicolinearidade envolvia as variáveis calçada e categoria. Assim, optou-se por retirar a variável calçada, uma vez que a categoria havia sido mantida no modelo. O mesmo motivo conduziu à retirada da variável edificação em lugar da variável asfalto. O critério para a retirada das variáveis água ou luz obedeceu ao mesmo princípio utilizado para Araçariguama, ou seja, o valor do R^2 . Como este não sofreu alteração alguma, optou-se, então, por retirar a variável água. O modelo obtido é mostrado na Tabela 5.14.

Tabela 5.10 - Melhores modelos de regressão ajustados, considerando o preço/m² para São Carlos.

	MODELO	R²
1	(Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água)	0,34
2	(Preço/m ²) = função (área ^{0.5} , e ^{testada*0.02} , forma, muro, calçada, e ^{-mc*0.2} , situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água)	0,38
3	(Preço/m ²) = função (área ^{0.5} , e ^{testada*0.02} , forma, muro, calçada, e ^{-mc*0.2} , situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*testada, área*mc, testada*mc)	0,38
4	(Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*testada, área*mc, testada*mc)	0,34
5	(Preço/m ²) = função (área ^{0.5} , testada ^{0.5} , forma, muro, calçada, mc ^{-0.2} , situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*testada, área*mc, testada*mc)	0,31
6	(Preço/m ²) = função (1/área, 1/testada, forma, muro, calçada, 1/mc, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água)	0,31
7	(Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*testada, área*mc, testada*mc, área*testada*mc)	0,37
8	Ln (Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*testada, área*mc, testada*mc)	0,51
9	Ln (Preço/m ²) = função (1/área, 1/testada, forma, muro, calçada, 1/mc, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*test, área*mc, testada*mc)	0,38
10	Ln (Preço/m ²) = função (1/área, 1/testada, forma, muro, calçada, 1/mc, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, área*test, área*mc, testada*mc, área*testada*mc)	0,51
11	(Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, categoria)	0,53
12	Ln (Preço/m ²) = função (área, testada, forma, muro, calçada, menor caminho, situação, topologia, asfalto, existência de edificações na quadra, esgoto, luz, água, categoria)	0,73

	ÁGUA	LUZ	ESG.	EDIF.	ASF.	CALÇ.	MURO	CAT.	SIT.	FORMA	TEST.	MC	ÁREA	TOP.
TOPOG.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,03	0,21	0,00
ÁREA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,81	0,10	0,66	0,00
MC	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,00	0,04	0,33	0,00	0,00
TESTADA	0,00	0,00	0,00	0,00	0,00	0,01	0,04	0,57	0,00	0,00	0,00	0,28	0,00	0,01
FORMA	0,00	0,00	0,00	0,00	0,00	0,22	0,21	0,02	0,03	0,01	0,00	0,04	0,00	0,01
SIT.	0,00	0,00	0,00	0,02	0,04	0,01	0,39	0,10	0,00	0,00	0,00	0,05	0,00	0,03
CATEG.	0,00	0,00	0,00	0,02	0,04	0,68	0,24	0,00	0,03	0,01	0,01	0,01	0,01	0,01
MURO	0,00	0,00	0,01	0,01	0,03	0,03	0,10	0,10	0,44	0,03	0,00	0,11	0,00	0,03
CALÇ.	0,00	0,00	0,00	0,00	0,21	0,02	0,00	0,08	0,39	0,22	0,02	0,00	0,01	0,01
ASF.	0,01	0,01	0,03	0,03	0,07	0,00	0,00	0,10	0,06	0,62	0,02	0,01	0,07	0,00
EDIF.	0,00	0,00	0,02	0,69	0,58	0,00	0,01	0,01	0,03	0,02	0,02	0,03	0,00	0,00
ESG.	0,00	0,01	0,93	0,23	0,03	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00
LUZ	0,02	0,07	0,00	0,00	0,00	0,00	0,00	0,01	0,03	0,06	0,01	0,01	0,00	0,00
ÁGUA	0,98	0,92	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00

Tabela 5.11 - Valores de π_{ij} obtidos para as colunas da matriz XX para São Carlos.

	TESTADA	ÁGUA	LUZ	ESG.	EDIF	ASF.	CALÇ.	MURO	CAT.	SIT.	FORMA	MC	ÁREA	TOPOG
TOPOG.	0,00	0,00	0,00	0,00	0,01	0,01	0,03	0,00	0,00	0,00	0,01	0,03	0,00	1,00
ÁREA	0,87	0,00	0,00	0,00	0,01	0,01	0,01	0,00	0,07	0,04	0,13	0,07	1,00	
MC	0,03	0,00	0,00	0,00	0,01	0,01	0,09	0,03	0,45	0,00	0,00	1,00		
FORMA	0,09	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,12	1,00			
SIT.	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00				
CAT.	0,05	0,00	0,00	0,00	0,00	0,03	0,06	0,02	1,00					
MURO	0,00	0,00	0,00	0,02	0,03	0,03	0,11	1,00						
CALÇ.	0,01	0,00	0,00	0,01	0,03	0,03	1,00							
ASF.	0,01	0,04	0,03	0,29	0,36	1,00								
EDIF.	0,01	0,05	0,04	0,33	1,00									
ESG.	0,00	0,11	0,09	1,00										
LUZ	0,00	0,80	1,00											
ÁGUA	0,00	1,00												
TESTADA	1,00													

Tabela 5.12 - Valores dos R_{ij}^2 obtidos para cada par de variáveis para a cidade de São Carlos.

Tabela 5.13 - Modelo de Regressão *Ridge* obtido para São Carlos utilizando o valor do m^2 como variável dependente.

	PARÂMETRO ESTIMADO	DESVIO-PADRÃO	ESTATÍSTICA t	p-level
CONSTANTE	4,312927	0,15962	27,01994	0
TOPOGRAFIA	0,052628	0,032766	1,606169	0,108917
AREA	-6,8E-05	0,000134	-0,50649	0,612753
MC	-0,07765	0,010264	-7,5657	2,09E-13
TESTADA	0,006377	0,00519	1,228618	0,219838
FORMA	0,029272	0,058809	0,497743	0,618901
SITUAÇÃO	-0,07487	0,042791	-1,74964	0,080842
TERR_B	-0,43542	0,034308	-12,6913	7,01E-32
TERR_C	-0,80181	0,044431	-18,0461	0
MURO	-0,04215	0,029719	-1,41819	0,156807
CALÇADA	0,008085	0,031415	0,257368	0,797008
ASFALTO	-0,0169	0,05512	-0,30654	0,759332
EDIFICAÇÃO	-0,01243	0,064541	-0,19259	0,847367
ESGOTO	0,117046	0,083914	1,394834	0,163733
LUZ	0,042453	0,262258	0,161876	0,871474
ÁGUA	0,037644	0,297839	0,126391	0,899477

Tabela 5.14 - Modelo de Regressão obtido para São Carlos utilizando o valor do m^2 como variável dependente após a retirada das variáveis testada, calçada, edificação e água.

	PARÂMETRO ESTIMADO	DESVIO-PADRÃO	ESTATÍSTICA t	p-level
CONSTANTE	4,277913	0,146914	32,52175	0
TOPOGRAFIA	0,051465	0,03197	1,60977	0,108122
ÁREA	8,32E-05	5,29E-05	1,571414	0,116762
MC	-0,07623	0,009909	-7,69263	8,6E-14
FORMA	0,040086	0,057437	0,697917	0,485576
SITUAÇÃO	-0,08141	0,042281	-1,92556	0,054765
TERR_B	-0,43542	0,034308	-12,6913	7,01E-32
TERR_C	-0,80181	0,044431	-18,0461	0
MURO	-0,04342	0,028126	-1,54361	0,123358
ASFALTO	-0,02148	0,050163	-0,42826	0,668662
ESGOTO	0,116266	0,077692	1,496499	0,135198
LUZ	0,061065	0,126934	0,48108	0,630684

Uma comparação entre as Tabelas 5.13 e 5.14 revela que não existe grande diferença entre o modelo estimado através de regressão *Ridge* e aquele encontrado após a retirada de algumas variáveis. Além do valor dos parâmetros,

uma comparação entre as somas dos Quadrados dos Resíduos (SSE) para os dois casos também foi feita. Para a regressão *Ridge*, a SSE obtida foi igual a 34,138 contra um valor igual a 33,659 para a outra alternativa. Desta forma, mais uma vez, devido à maior confiabilidade com relação aos testes de hipóteses, optou-se por utilizar o modelo apresentado na Tabela 5.14.

Antes dos testes t, porém, as suposições sobre a distribuição e variância do distúrbio aleatório deviam ser verificadas. Os testes utilizados para verificação destas suposições foram os mesmos do caso anterior, ou seja, Qui-Quadrado para verificar Normalidade e *Teste Geral de White* e *Teste de Goldfeld-Quandt* para homocedasticidade.

A estatística obtida para o teste de Normalidade foi igual a 65,89, cujo valor foi comparado ao tabelado da Distribuição Qui-Quadrado para um valor crítico de 95%, donde a suposição de distribuição Normal não foi rejeitada.

Para analisar a suposição de heterocedasticidade foi elaborado, a princípio, o *Teste Geral de White*. A estatística de teste obtida, com valor igual a 4,35, foi comparada com o valor tabelado da Distribuição Qui-quadrado para um valor crítico igual a 95% e 10 graus de liberdade (18,31). Como o valor calculado ficou menor que o tabelado, a hipótese de homocedasticidade não foi rejeitada.

Para o *Teste de Goldfeld-Quandt*, a amostra foi ordenada de acordo com o valor da variável área e dividida em três grupos, o primeiro e o último com 168 pontos e o intermediário com 144 pontos. A hipótese de que a variância no primeiro grupo seria maior que a do último foi rejeitada a um nível de significância igual a 5% (estatística de teste obtida foi igual a 0,324 - cujo valor foi comparado ao valor tabelado da Distribuição F para um valor crítico igual a 95% com 158 e 158 graus de liberdade). Portanto, por este teste, a hipótese de homocedasticidade também não foi rejeitada.

Obedecidas as suposições de Normalidade e variância constante, os testes t com relação aos parâmetros do modelo puderam, então, ser executados. Os valores mostrados na Tabela 5.14 indicam que os únicos parâmetros significativos são os estimados para as variáveis mc e categoria. O sinal negativo observado para o coeficiente da variável mc indica que um aumento na distância ao centro provoca

uma diminuição no preço. Através dos resultados, pode-se notar também que os logaritmos dos preços por m² de lotes situados nas categorias B e C são diminuídos de 0,435 e 0,870, respectivamente, com relação àqueles situados na categoria A.

5.2.2. APLICAÇÃO DAS REDES NEURAIS ARTIFICIAIS PARA SÃO CARLOS

Da mesma forma que para o estudo de caso anterior, para o emprego das RN, a princípio, o conjunto de dados foi dividido de forma aleatória em três subconjuntos distintos: treinamento (50% dos casos), validação (25% dos casos) e teste (25% dos casos). Este procedimento foi então repetido três vezes, o que gerou três diferentes conjuntos de dados. Redes com vários tipos de configurações foram então simuladas e o erro relativo total fornecido por cada uma destas, calculado. A taxa de aprendizado, o *momentum* e o erro admitido para a fase de treinamento também foram variados. Inicialmente, 13 variáveis (área, testada, esquina, muro, calçada, existência de edificações na quadra, distância ao centro, topografia, água, luz, esgoto, asfalto e forma) foram utilizadas. Como os resultados não foram favoráveis, uma décima quarta variável, indicando a categoria em que o terreno se enquadra (A, B ou C), de acordo com as categorias de amostragem, foi adicionada. Os resultados da Tabela 5.15, obtidos para um erro admitido na fase de treinamento igual a 0,03, mostram a média dos erros obtidos pela utilização dos três conjuntos de dados, assim como as topologias das redes testadas quando o valor do m² foi utilizado como variável resposta.

Tabela 5.15 - Resultados das melhores simulações, usando todas as variáveis e o preço do m² para São Carlos.

NÓS NA CAMADA INTERM.	TAXA DE APREN.	MOMENTUM	MÉDIA DOS ERROS	R ² (conj. 1)	R ² (conj. 2)	R ² (conj. 3)
5	0,5	0,8	30,08	0,51	0,40	0,50
6	0,3	0,4	28,99	0,53	0,48	0,55
6	0,5	0,8	31,11	0,27	0,32	0,53
13	0,5	0,8	29,26	0,54	0,43	0,47
6 - 4	0,5	0,8	27,88	0,53	0,41	0,55
13 - 5	0,5	0,8	27,87	0,52	0,42	0,56
14 - 6	0,5	0,8	27,63	0,53	0,44	0,56

Conforme pode ser observado na Tabela 5.15, os valores obtidos para o R^2 ficaram baixos com relação ao observado para o modelo de regressão. Desta forma, foi feita uma tentativa de transformação logarítmica na variável resposta com o objetivo de verificar se a saída fornecida pela rede iria se adequar melhor aos dados. Os novos resultados obtidos, para um erro de treinamento igual a 0,03, são mostrados na Tabela 5.16.

Tabela 5.16 - Resultados das melhores simulações, usando todas as variáveis e o ln do preço do m^2 para São Carlos.

NÓS NA CAMADA INTERM.	TAXA DE APREN.	<i>MOMENTUM</i>	MÉDIA DOS ERROS	R^2 (conj. 1)	R^2 (conj. 2)	R^2 (conj. 3)
6	0,5	0,6	7,86	0,63	0,56	0,65
6	0,5	0,8	7,91	0,63	0,54	0,65
6	0,4	0,4	7,74	0,69	0,64	0,73
6	0,3	0,4	7,86	0,68	0,56	0,73
6	0,5	0,5	7,81	0,68	0,56	0,72
6	0,4	0,3	7,31	0,68	0,66	0,73
6	0,4	0,2	7,74	0,68	0,68	0,73
9 - 8	0,5	0,8	7,38	0,66	0,64	0,66
10 - 8	0,5	0,8	7,41	0,63	0,69	0,68
14 - 6	0,5	0,8	7,14	0,65	0,67	0,66
8 - 7	0,5	0,8	7,14	0,66	0,66	0,69
8 - 6	0,5	0,8	7,70	0,64	0,66	0,68
8 - 8	0,5	0,8	6,97	0,67	0,63	0,66
8 - 9	0,5	0,8	7,60	0,65	0,66	0,68
9 - 8	0,4	0,4	7,90	0,65	0,66	0,69
8 - 7	0,4	0,4	7,40	0,67	0,63	0,67

Os valores obtidos para o R^2 observados na Tabela 5.16 mostraram uma melhora significativa quando a transformação logarítmica foi aplicada à variável resposta. Pode-se observar também que a melhor configuração de rede observada foi aquela com 6 nós na camada intermediária, taxa de aprendizado igual a 0,4 e *momentum* igual a 0,2.

Da mesma forma que o caso anterior, a importância de cada variável no valor estimado final foi calculada através do método proposto por Garson. Os resultados obtidos são mostrados na Tabela 5.17.

Tabela 5.17 - Relevância de cada variável para a cidade de São Carlos.

CONJUNTO 1		CONJUNTO 2		CONJUNTO 3	
VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)
CATEGORIA	17	MC	17	MC	16
MC	15	CATEGORIA	14	CATEGORIA	14
MURO	12,5	MURO	7,5	MURO	12
ÁREA	9,5	ÁREA	7,5	ÁREA	11
ÁGUA	7	ESGOTO	7,5	ÁGUA	8
LUZ	6	CALÇADA	6,5	CALÇADA	7
ESGOTO	6	FORMA	6,5	TESTADA	5
CALÇADA	5,5	EDIFICAÇÃO	6	EDIFICAÇÃO	5
SITUAÇÃO	4	LUZ	6	LUZ	5
TESTADA	4	ÁGUA	5	ESGOTO	4
EDIFICAÇÃO	3,5	TOPOGRAFIA	5	SITUAÇÃO	3,7
TOPOGRAFIA	3,5	SITUAÇÃO	4	TOPOGRAFIA	3,7
ASFALTO,	3	ASFALTO	3,6	FORMA	2,5
FORMA	3	TESTADA	3,6	ASFALTO	1,5

Os dados observados na Tabela 5.17 revelam que, para a cidade de São Carlos, a variável mc aparece em primeiro lugar em dois dos conjuntos de dados, seguida pela categoria a que o lote pertence. No primeiro conjunto, contudo, a distância ao centro aparece em segundo lugar, mas também com uma contribuição significativa para o valor final, igual a 15%. Observa-se que a existência de muro representa uma variável importante na valorização de lotes, uma vez que aparece em terceiro lugar para todos os conjuntos de dados. A variável área em nenhum dos casos apresentou contribuição superior a 11%. Da mesma forma que para o modelo de regressão, a existência de asfalto não causa influência significativa no preço de terrenos, fato este que pode ser explicado pelo alto nível de pavimentação da cidade. Devido ao tipo de relevo acidentado comum na cidade, lotes com topografia irregular também não são desvalorizados. A variável testada também não aparece como fator relevante.

Conforme proposto na Metodologia, as variáveis cuja influência fosse igual ou inferior a 5% seriam retiradas do conjunto de dados e a rede, novamente testada. Desta forma, as variáveis topografia, situação, testada e asfalto foram retiradas do conjunto de dados e a rede novamente treinada. Esta nova situação, porém, não gerou resultados estatisticamente significativos para um nível crítico

igual a 95%, uma vez que o erro médio observado ficou igual a 7,71 e os R^2 obtidos foram 0,69, 0,68 e 0,73 para os conjuntos 1, 2 e 3, respectivamente.

5.2.3. COMPARAÇÃO DOS RESULTADOS OBTIDOS PELOS DOIS MÉTODOS PARA A CIDADE DE SÃO CARLOS

Os mesmos conjuntos de dados utilizados para a fase de teste das Redes Neurais foram utilizados para estimar os erros e R^2 obtidos pelo modelo de regressão. Estes resultados são mostrados na Tabela 5.18.

Tabela 5.18 - Erros e R^2 obtidos para os dois modelos e os três conjuntos de dados para São Carlos.

	REGRESSÃO		RN	
	ERRO	R^2	ERRO	R^2
conjunto 1	6,38	0,70	6,72	0,69
conjunto 2	7,93	0,67	7,80	0,68
conjunto 3	8,55	0,70	8,62	0,73
MÉDIA	7,62	-	7,71	-

Os resultados mostrados na Tabela 5.18 indicam que o erro relativo total obtido pelas RN ficou em torno de 1,2% maior que aquele fornecido pela regressão. Quanto ao valor do R^2 , pode-se observar que o segundo método mostrou uma ligeira superioridade para os conjuntos 2 e 3. Desta forma, os resultados obtidos até aqui indicam uma certa coerência em termos de erros de estimação e R^2 para os dois métodos, não se podendo notar mudanças significativas quando um ou outro foi utilizado.

O próximo passo constou em comparar os erros relativos para os dois métodos, através dos gráficos mostrados nas Figuras 5.10, 5.11 e 5.12, mostrando estes valores para ambos os métodos e os conjuntos de dados 1, 2 e 3, respectivamente.

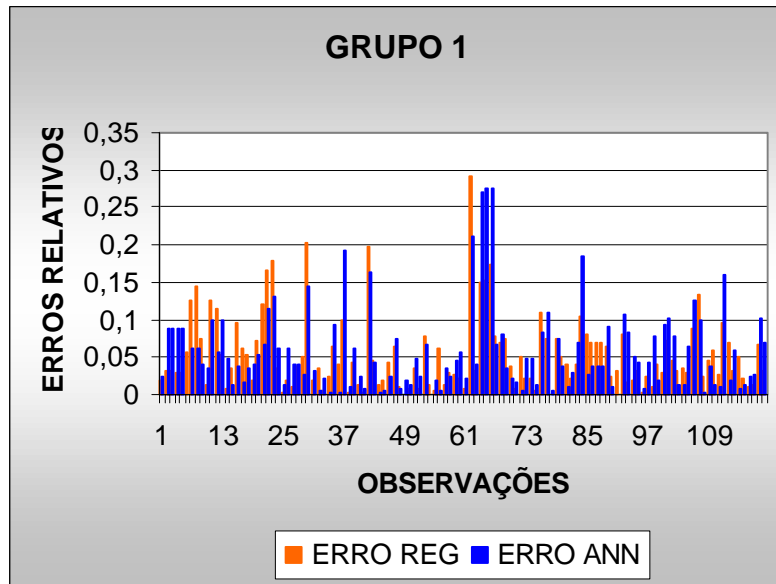


Figura 5.10 - Erros relativos obtidos para os dois métodos e o primeiro conjunto de dados para a cidade de São Carlos.

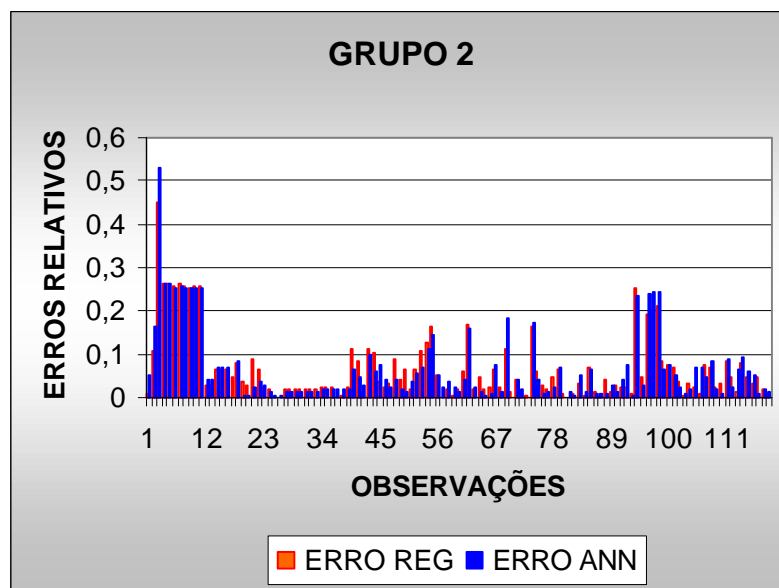


Figura 5.11 - Erros relativos obtidos para os dois métodos e o segundo conjunto de dados para a cidade de São Carlos.

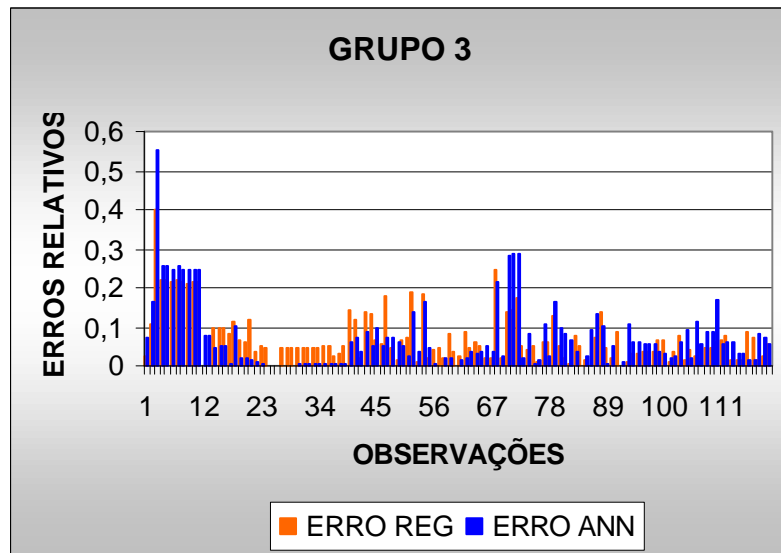


Figura 5.12 - Erros relativos obtidos para os dois métodos e o terceiro conjunto de dados para a cidade de São Carlos.

As Figuras 5.10, 5.11 e 5.12 revelam um comportamento bastante parecido, em termos de erro de estimação. Pode-se notar também que, para os três casos, quando o valor do erro relativo atingiu seu pico para um dado modelo, o mesmo ocorreu para o outro, o que pode ser indicativo da existência de um ponto cujo valor observado era muito discrepante do restante do conjunto de dados.

O próximo passo constou do agrupamento dos valores dos erros em *clusters*, visando a construção de mapas temáticos e posterior identificação de um possível padrão espacial nos valores dos erros. A divisão em *clusters*, foi obtida através do emprego do mesmo método utilizado para Araçariçuama e é mostrada na Tabela 5.19.

Tabela 5.19 - Proporção de erros relativos nos *clusters* encontrados com os dois modelos para os três conjuntos de dados para São Carlos.

Conjunto/ <i>Cluster</i>	Limites de classe	REGRESSÃO	RN
Conjunto 1 - <i>Cluster</i> 1	$\leq 0,0267$	35,0%	36,0%
Conjunto 1 - <i>Cluster</i> 2	0,0268 a 0,0638	35,0%	31,0%
Conjunto 1 - <i>Cluster</i> 3	0,0639 a 0,1327	22,5%	24,0%
Conjunto 1 - <i>Cluster</i> 4	$\geq 0,1328$	8,5%	7,5%
Conjunto 2 - <i>Cluster</i> 1	$\leq 0,0381$	53,0%	52,0%
Conjunto 2 - <i>Cluster</i> 2	0,0382 a 0,0880	27,5%	30,0%
Conjunto 2 - <i>Cluster</i> 3	0,0881 a 0,1989	10,0%	7,5%
Conjunto 2 - <i>Cluster</i> 4	$\geq 0,1990$	9,5%	10,5%
Conjunto 3 - <i>Cluster</i> 1	$\leq 0,0355$	33,0%	42,5%
Conjunto 3 - <i>Cluster</i> 2	0,0355 a 0,0891	44,0%	34,0%
Conjunto 3 - <i>Cluster</i> 3	0,0892 a 0,1725	12,5%	12,5%
Conjunto 3 - <i>Cluster</i> 4	$\geq 0,1726$	11,0%	11,0%

Diferentemente do que aconteceu para Araçariçuama, não se percebe uma grande diferença na concentração de pontos situados dentro dos intervalos que apresentaram valores de erros menores (*clusters* 1 e 2) para os dois métodos. Portanto, parece não haver diferença entre regressão e Redes Neurais quando a análise de *clusters* foi utilizada.

Uma análise dos mapas temáticos mostrados nas Figuras 5.13, 5.14, 5.15, 5.16, 5.17 e 5.18 mostra que para os elementos do grupo 1 parece não haver padrão espacial definido para a distribuição dos erros tanto para um método quanto para outro. Para os grupos 2 e 3, no entanto, nota-se uma concentração de erros maiores no bairro situado no canto inferior esquerdo para ambos os modelos. As Figuras 5.15, 5.16, 5.17 e 5.18 revelam uma grande semelhança no padrão de distribuição espacial para os erros obtidos para os métodos de regressão e Redes Neurais.

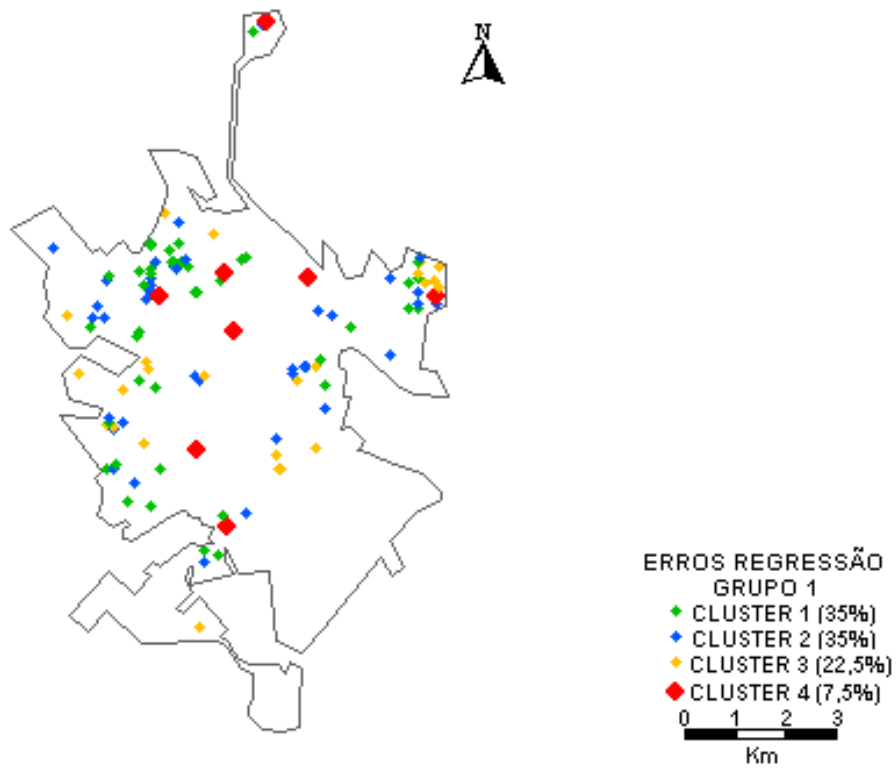


Figura 5.13 - Distribuição espacial dos erros para o modelo de regressão e o primeiro grupo de dados de São Carlos.

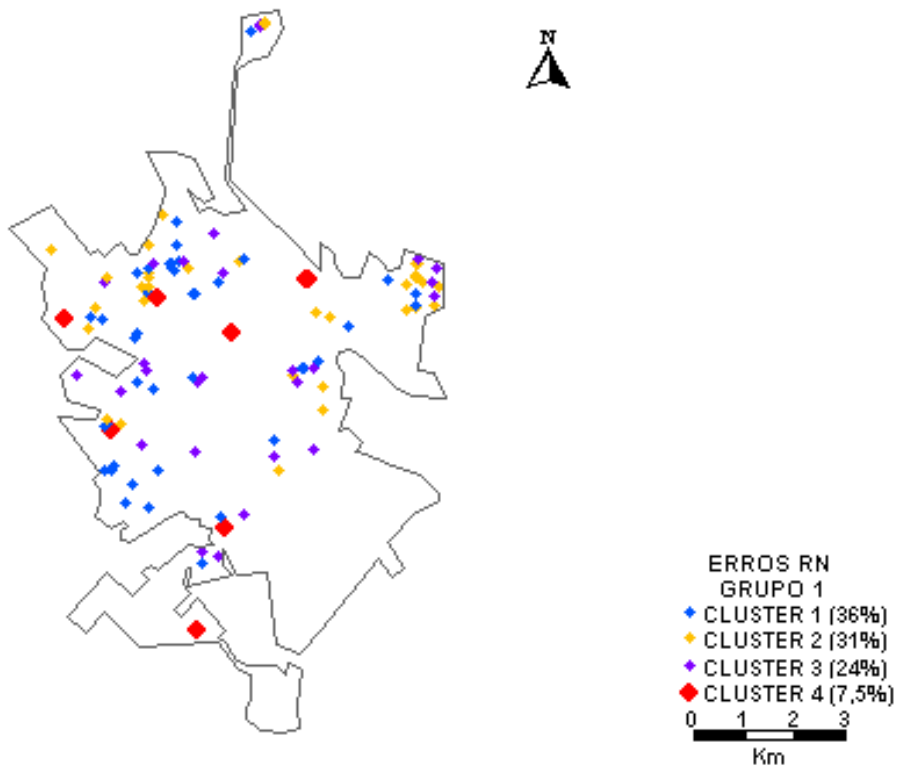


Figura 5.14 - Distribuição espacial dos erros para as RN e o primeiro grupo de dados de São Carlos.

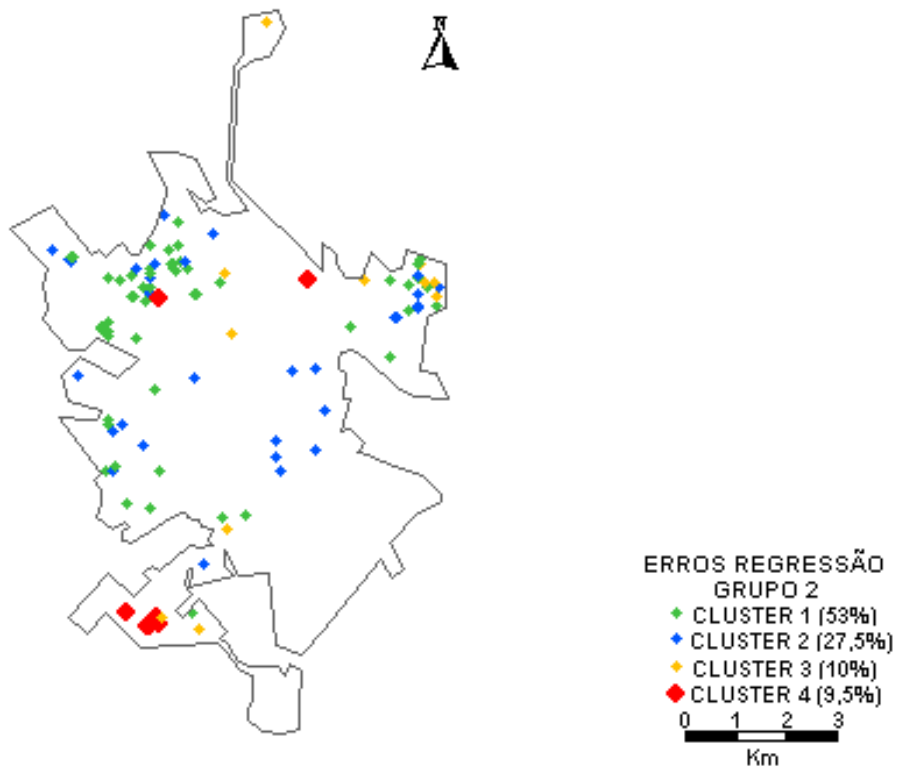


Figura 5.15 - Distribuição espacial dos erros para o modelo de regressão e o segundo grupo de dados de São Carlos.

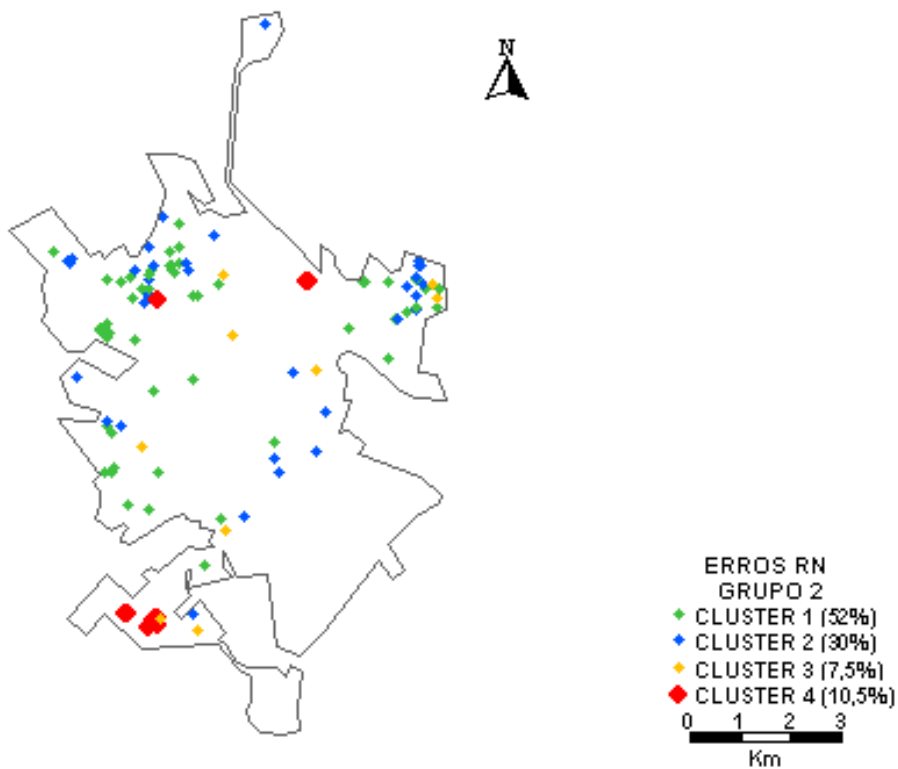


Figura 5.16 - Distribuição espacial dos erros para as RN e o segundo grupo de dados de São Carlos.

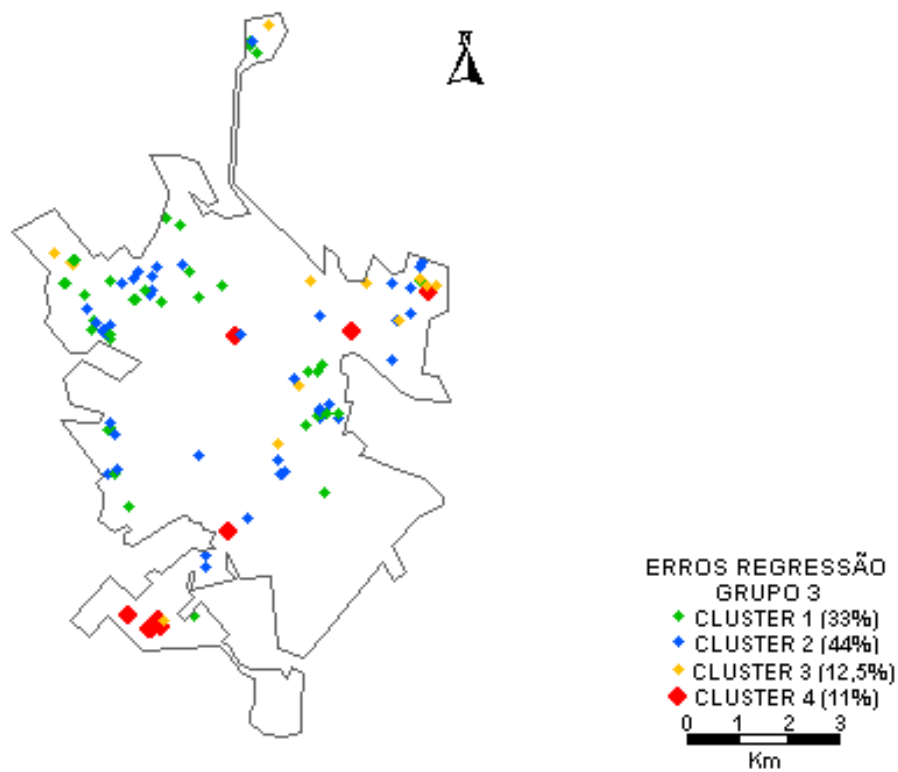


Figura 5.17 - Distribuição espacial dos erros para o modelo de regressão e o terceiro grupo de dados de São Carlos.

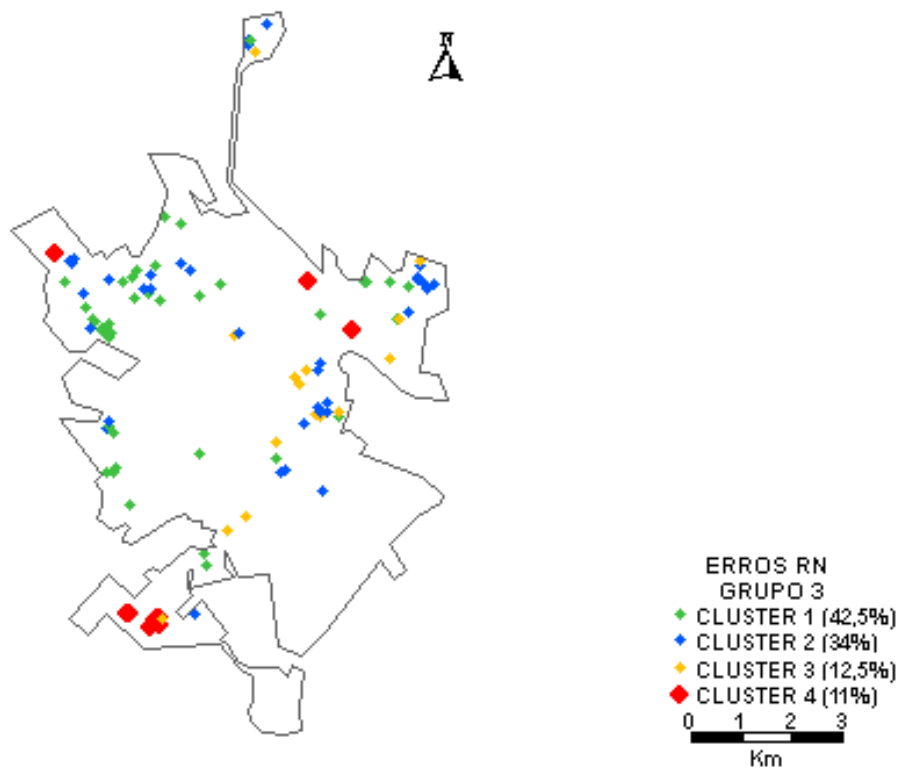


Figura 5.18 - Distribuição espacial dos erros para as RN e o terceiro grupo de dados de São Carlos.

5.3. AVALIAÇÃO DO DESEMPENHO DAS REDES NEURAS

Uma das principais características de uma Rede Neural é a de que ela é capaz de fornecer boas estimativas mesmo quando as variáveis independentes possuem algum tipo de relação desconhecida entre si. Esta é uma das vantagens do método com relação aos modelos de estimação convencionais, uma vez que nem sempre é possível saber com exatidão que tipo de relação envolve duas ou mais variáveis de entrada. O mesmo se aplica com respeito às variáveis de saída e de entrada, uma vez que o esperado é que a rede seja capaz de captar o tipo de relação existente entre elas e fornecer estimativas adequadas.

Para os estudos de caso mostrados aqui, a rede foi treinada utilizando o valor m^2 ou uma função deste como variável resposta. Porém, tornou-se interessante verificar como as RN reagiriam se uma nova situação fosse apresentada, onde o valor total do terreno fosse utilizado como entrada.

A Tabela 5.20 mostra os resultados obtidos para a cidade de Araçariguama, quando o valor total do terreno foi utilizado na saída.

Tabela 5.20 – Resultados obtidos pelas Redes Neurais quando o valor total do terreno foi utilizado como saída para a cidade de Araçariçuama.

	Número de nós nas camadas intermediárias	Taxa de aprendizado	Momentum	Média dos erros relativos
Erro de treinamento = 0,05				
Dados brutos	4	0,5	0,8	24,96
Dados brutos	6	0,5	0,8	23,87
Dados brutos	4	0,3	0,4	24,08
Dados brutos	2	0,4	0,4	24,82
Dados brutos	6	0,4	0,4	25,16
Dados brutos	3	0,2	0,1	34,78
Dados brutos	3	0,2	0,1	24,36
Norm. 0 - 1	6	0,2	0,7	23,34
Norm. 0 - 1	4	0,3	0,4	23,64
Norm. 0 - 1	7 - 6	0,3	0,4	23,69
Norm. 0 - 1	5 - 3 - 2	0,5	0,8	23,40
Norm. 0 - 1	3	0,3	0,2	23,79
Norm. 0 - 1	6-5-4-2	0,3	0,4	23,92
Erro de treinamento = 0,01				
Dados brutos	4	0,5	0,8	27,90
Dados brutos	6	0,5	0,8	29,09
Dados brutos	4-3	0,5	0,8	50,65
Dados brutos	6-3	0,1	0,2	27,20
Dados brutos	7	0,5	0,6	23,57
Norm. 0 - 1	7	0,5	0,8	27,17
Norm. 0 - 1	4	0,5	0,8	36,73
Norm. 0 - 1	4	0,3	0,4	26,95

Os valores observados na Tabela 5.20 indicam que a rede com 4 nós na camada intermediária, taxa de aprendizado igual a 0,3 e *momentum* igual a 0,4 apresentou erro relativo médio situado entre os menores valores observados, independentemente do erro de treinamento (0,05 ou 0,01). Vale observar que para o caso onde se usou o m^2 como variável resposta, a configuração de rede escolhida foi a mesma. Portanto, estes resultados indicam que, independentemente da transformação a que se sujeitou os dados, a configuração de rede citada forneceu bons resultados. Uma análise da saliência das variáveis também foi executada para este caso, cujos resultados são apresentados na Tabela 5.21.

Tabela 5.21 - Relevância de cada variável para a cidade de Araçariquama quando o preço total do terreno foi utilizado .

CONJUNTO 1		CONJUNTO 2		CONJUNTO 3	
VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)
ÁREA	16	ÁREA	16	FORMA	17
FORMA	16	FORMA	16	ÁREA	16
SITUAÇÃO	16	SITUAÇÃO	15	SITUAÇÃO	16
TESTADA	15	TESTADA	15	TESTADA	15
MC	13	MC	14	MC	13
CALÇADA	13	CALÇADA	12	MURO	11,5
MURO	11	MURO	11	CALÇADA	11,5

Os valores observados na Tabela 5.21 indicam que a variável mc aparece em quinto lugar na lista de importâncias relativas, sempre precedido pelas variáveis área, testada, forma e situação. Portanto, quando o preço total do lote está sendo estimado pelas Redes Neurais, suas características físicas são as mais consideradas.

O mesmo tipo de análise foi elaborado para a cidade de São Carlos, e os resultados são mostrados na Tabela 5.22.

Tabela 5.22 - Resultados obtidos pelas Redes Neurais quando o valor total do terreno foi utilizado como saída para a cidade de São Carlos.

NÚMERO NÓS NA ENTRADA	NÓS NA CAMADA INTERME- DIÁRIA	TAXA DE APREN.	<i>MOMENTUM</i>	MÉDIA DOS ERROS
ERRO ADMITIDO PARA A FASE DE TREINAMENTO = 0,05				
14	6	0,5	0,8	139,103
14	8 - 5	0,5	0,8	156,997
14	3	0,5	0,8	140,333
14	6	0,3	0,4	156,767
14	8 - 5	0,3	0,4	162,793
14	3	0,3	0,3	152,293
13	6	0,5	0,8	138,327
13	11	0,5	0,8	134,327
13	5 - 3	0,5	0,8	163,123
13	5 - 3	0,3	0,3	163,777
13	11	0,3	0,3	161,517
ZONA 1: VAR 6 A 13	3	0,5	0,8	151,253
ZONA 1: VAR 1 E 2 ZONA 2: VAR 6 A 13	3	0,3	0,3	156,470
ZONA 1: VAR 1 E 2 ZONA 2: VAR 6 A 13	6	0,5	0,8	151,493

ZONA 1: VAR 1 E 2 ZONA 2: VAR 6 A 13	6	0,3	0,3	164,657
13	6 - 5 - 4	0,3	0,3	166,330
13	4 - 3	0,2	0,1	173,063
13	7 - 6 - 4	0,2	0,2	189,657
13	12 - 11	0,3	0,3	165,383
ERRO ADMITIDO PARA A FASE DE TREINAMENTO = 0,01				
13	5 - 3	0,3	0,3	145,413
14	6	0,3	0,4	53,453
14	8 - 5	0,5	0,8	89,357
14	3	0,5	0,8	57,520
14	5 - 3	0,5	0,8	102,563
13	6	0,5	0,8	68,610
13	11	0,5	0,8	65,147
ZONA 1: VAR 6 A 13	3	0,5	0,8	61,507
ZONA 1: VAR 1 E 2 ZONA 2: VAR 6 A 13	6	0,5	0,8	86,703
13	6 - 5 - 4	0,5	0,8	134,477
13	4 - 3	0,5	0,8	128,153
13	7 - 6 - 4	0,5	0,8	131,873
13	12 - 11	0,5	0,8	95,240
14	5	0,3	0,3	104,737
14	6 - 5	0,3	0,3	118,893
14	6	0,5	0,8	53,993
14	4	0,5	0,8	54,600
14	5 - 4	0,5	0,8	107,970

A Tabela 5.22 mostra os resultados obtidos para as redes simuladas para São Carlos com 13 e 14 nós na camada de entrada e erros de treinamento iguais a 0,05 e 0,01, onde a rede com 14 nós incluiu a variável categoria.

De acordo com os dados fornecidos, chegou-se à conclusão de que a rede que apresentou os melhores resultados foi aquela com 14 nós na camada de entrada, 6 na camada intermediária, taxa de aprendizado igual a 0,3 e *momentum* igual a 0,4, considerando o erro de treinamento igual a 0,01. Novamente, observou-se que esta configuração também conduziu a um dos menores erros médios obtidos para o modelo onde se usou o $\ln(m^2)$, ou seja, independentemente da transformação aplicada, esta configuração de rede conduziu a bons resultados. Uma análise da saliência das variáveis é mostrada na Tabela 5.23.

Tabela 5.23 - Relevância de cada variável para a cidade de São Carlos, quando o preço total do terreno foi considerado.

CONJUNTO 1		CONJUNTO 2		CONJUNTO 3	
VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)	VARIÁVEL	RELEV. (%)
ÁREA	23	ÁREA	24	ÁREA	31
CATEGORIA	12	CATEGORIA	12	MC	10
MC	8,5	MC	8	CATEGORIA	7
EX. EDIF.	8,5	EX. EDIF.	8	EX. EDIF	7
TESTADA	6,5	TESTADA	6,5	FORMA	6
FORMA	6,5	FORMA	6,5	SITUAÇÃO	6
ASFALTO	6	ASFALTO	6	ASFALTO	6
SITUAÇÃO	6	SITUAÇÃO	6	ESGOTO	5
ESGOTO	4,5	ESGOTO	4,5	TESTADA	5
LUZ	4,5	LUZ	4,5	LUZ	5
MURO	4	MURO	4	MURO	5
CALÇADA	4	CALÇADA	4	CALÇADA	4
TOPOGRAFIA	3	TOPOGRAFIA	3	TOPOGRAFIA	4
ÁGUA	3	ÁGUA	3	ÁGUA	3

Os dados mostrados na Tabela 5.23 indicam que a variável área apresenta a maior importância no valor estimado, como já era esperado, uma vez que o preço total do lote está sendo considerado. Quanto à variável mc, nota-se que esta aparece em terceiro lugar de importância para dois dos conjuntos de dados, sendo precedida pela variável categoria. Para o terceiro conjunto, no entanto, a acessibilidade aparece em segundo lugar na escala de relevâncias.

Os resultados observados nas Tabelas 5.20 e 5.22 indicam que, em se tratando de estimação de valor, as Redes Neurais apresentam uma certa coerência em termos de configuração, independentemente da transformação empregada nas variáveis. Assim, para Araçariguama, uma rede com 4 nós na camada intermediária, taxa de aprendizado igual a 0,4 e *momentum* igual a 0,3 seria adequada tanto para estimar o valor total dos lotes quanto o valor do m², dependendo do conjunto de dados de treinamento. O mesmo foi observado para São Carlos, onde uma rede com 6 nós na camada intermediária, taxa de aprendizado igual a 0,3 e *momentum* igual a 0,4 seria adequada tanto para estimar o valor total quanto unitário.

Porém, ao analisar a importância de cada variável na resposta fornecida para Araçariguama, percebe-se uma diferença em termos de importância, uma vez que a acessibilidade, que teve uma contribuição superior a 34% para a estimação do

valor do m^2 , se apresenta em quinto lugar para a nova situação, sempre precedida pela área e outras características físicas e com representação relativa variando de 13% a 15%.

Da mesma maneira que ocorreu para Araçariguama, os dados apresentados na Tabela 5.23 indicam que, para São Carlos, a variável mais relevante é a área, fato já esperado quando se considera o valor total do lote. Por sua vez, a acessibilidade aparece em terceiro lugar, precedida pela variável categoria e no último conjunto estas posições se invertem. No caso onde se analisou o preço do m^2 , contudo, verificou-se que a acessibilidade aparecia em primeiro lugar para dois dos conjuntos.

Diante dos resultados mostrados nas Tabelas 5.21 e 5.23, resolveu-se fazer uma comparação dos erros relativos obtidos para os dois casos (estimação do valor total e do m^2 para as duas cidades e os três conjuntos de dados). Os resultados obtidos para Araçariguama são mostrados nas Figuras 5.19 a 5.21 para os grupos 1, 2 e 3, respectivamente. Para São Carlos, a comparação gráfica destes erros é mostrada nas Figuras 5.22 a 5.24.

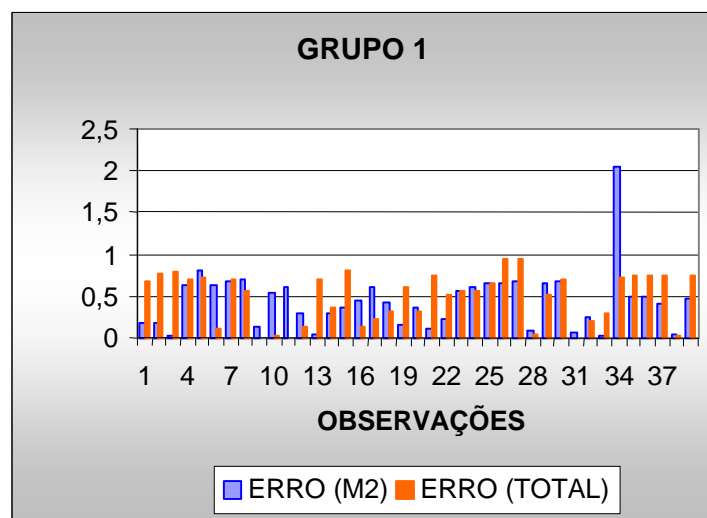


Figura 5.19 – Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o primeiro conjunto de dados e a cidade de Araçariguama.

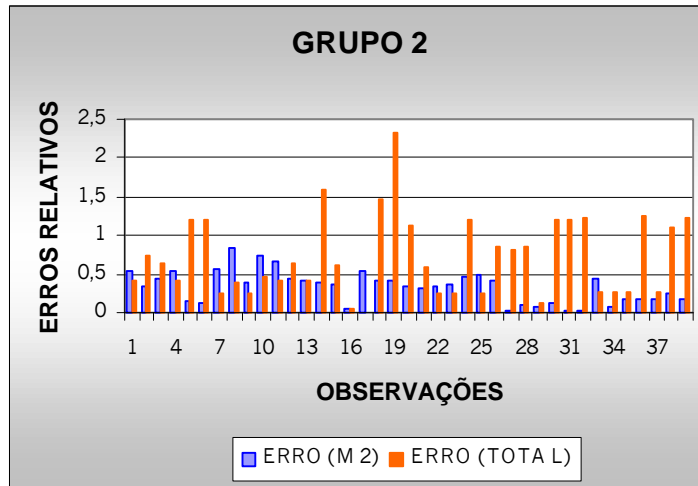


Figura 5.20 – Erros relativos obtidos pelas RN para o m² e o preço total do lote para o segundo conjunto de dados e a cidade de Araçariçuama.

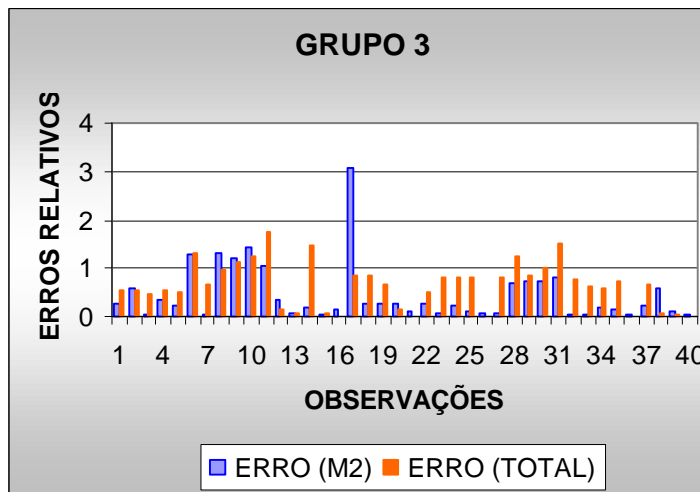


Figura 5.21 – Erros relativos obtidos pelas RN para o m² e o preço total do lote para o terceiro conjunto de dados e a cidade de Araçariçuama.

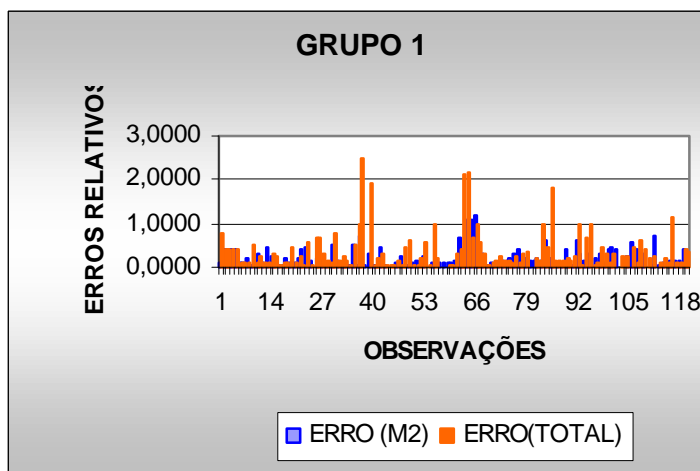


Figura 5.22 – Erros relativos obtidos pelas RN para o m² e o preço total do lote para o primeiro conjunto de dados e a cidade de São Carlos.

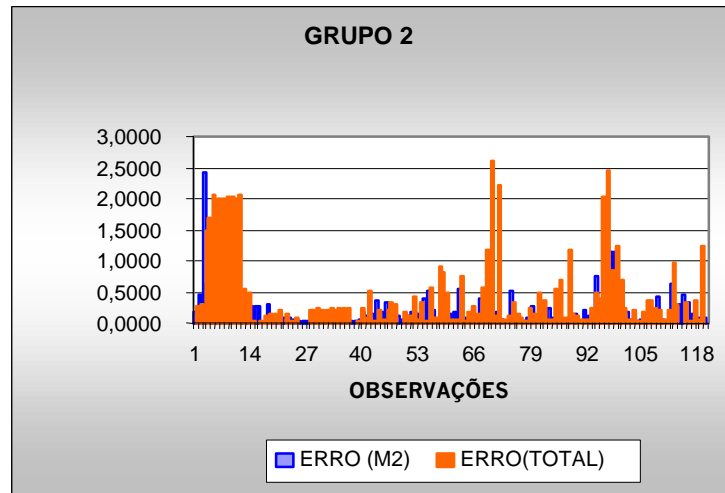


Figura 5.23 – Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o segundo conjunto de dados e a cidade de São Carlos.

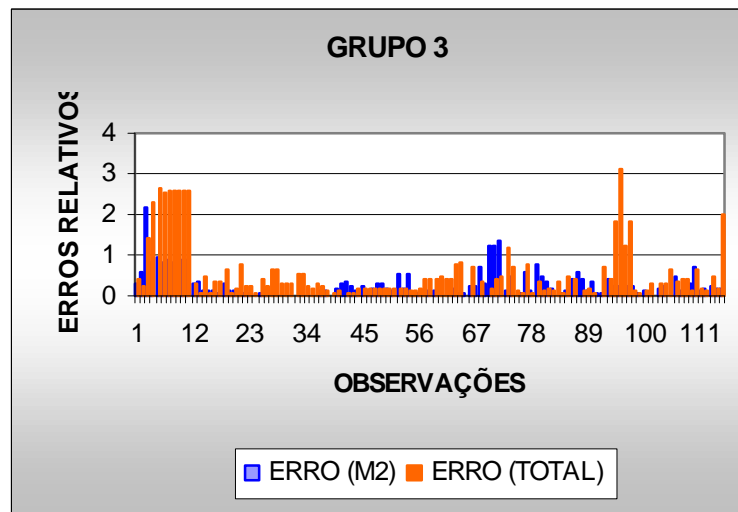


Figura 5.24 – Erros relativos obtidos pelas RN para o m^2 e o preço total do lote para o terceiro conjunto de dados e a cidade de São Carlos.

Os resultados observados nas Figuras 5.19 a 5.24 indicam que quando o valor do m^2 foi usado em lugar do valor total do terreno, no geral os erros relativos foram menores. Para a cidade de Araçariçuama, observa-se que para os conjuntos 1 e 3 há um pico no valor do erro quando o m^2 foi utilizado. Porém, para os demais casos, verifica-se que o erro obtido para o preço total é, na maioria das vezes, maior. Para São Carlos verifica-se que apenas para o grupo 2 existe um pico no erro relativo obtido para o modelo que utilizou o m^2 . Na maioria dos casos restantes, os erros obtidos pela utilização do valor total foram maiores.

Embora alguns autores (DANTAS, 1999; MENDONÇA et al., 1998) também utilizem o valor total do lote sem maiores prejuízos, optou-se aqui por basear as análises no valor do m^2 , uma vez que os erros fornecidos pela sua utilização foram menores que os obtidos pelo modelo que utilizou o valor total.

Os mapas temáticos mostrados nas Figuras 5.25 a 5.27 mostram a distribuição espacial dos erros fornecidos pelas RN quando o valor total do lote foi utilizado para a cidade de São Carlos. Uma comparação entre estes mapas e aqueles obtidos pela utilização do m^2 indica que a concentração de erros maiores apresenta aproximadamente o mesmo padrão, independentemente da transformação a que as variáveis foram sujeitas.

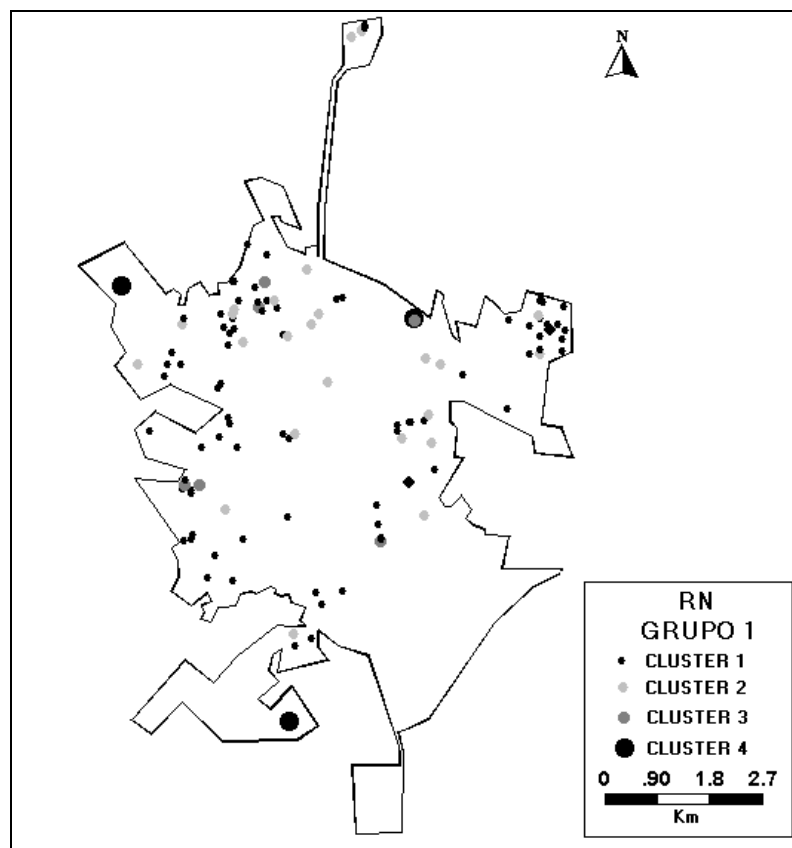


Figura 5.25 - Distribuição espacial dos erros obtidos para as RN e o primeiro grupo de dados de São Carlos quando o valor total do lote foi utilizado.

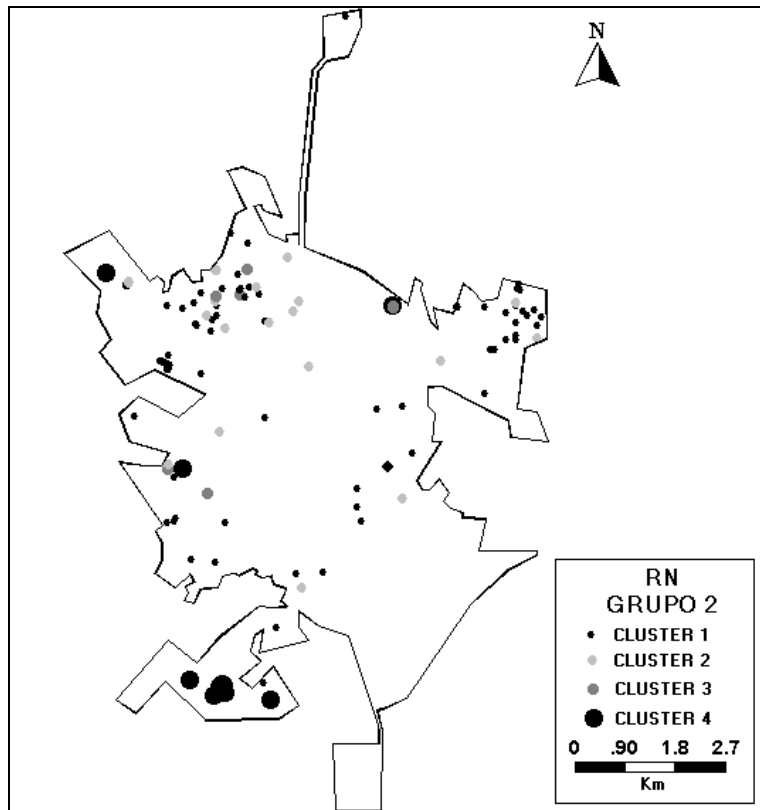


Figura 5.26 - Distribuição espacial dos erros obtidos para as RN e o segundo grupo de dados de São Carlos quando o valor total do lote foi utilizado.

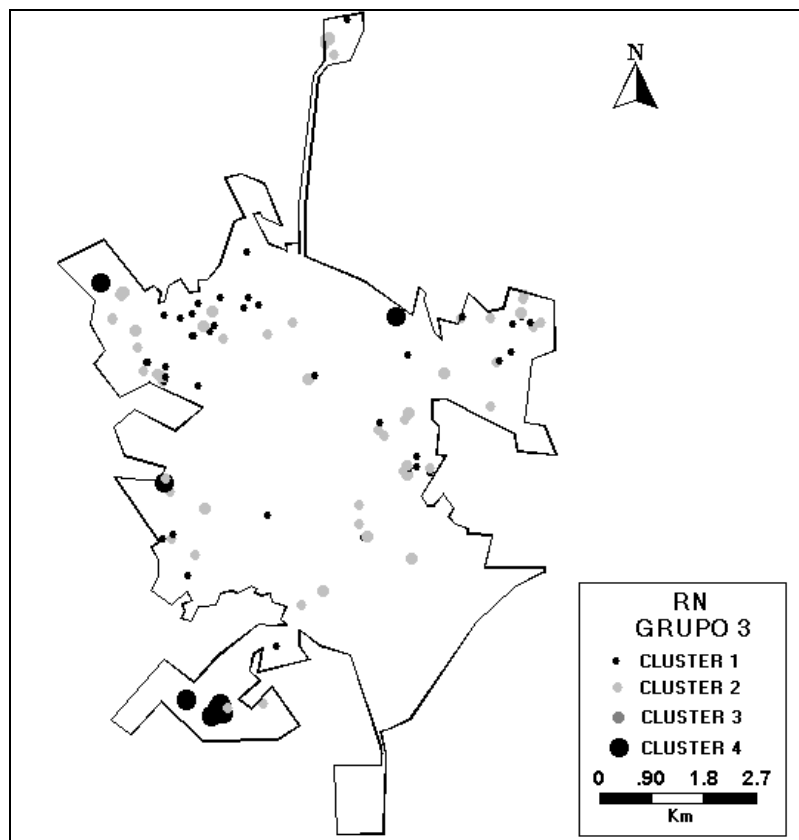


Figura 5.27 - Distribuição espacial dos erros obtidos para as RN e o terceiro grupo de dados de São Carlos quando o valor total do lote foi utilizado.

6. CONCLUSÕES

Este trabalho tinha três objetivos, a saber: identificar as variáveis que interferem no valor das propriedades, avaliar o uso de Redes Neurais Artificiais para fins de avaliação e estudar a influência de uma medida de acessibilidade no valor de terrenos urbanos. Os resultados obtidos serão comentados a seguir.

Embora não tenham sido observados muitos casos na literatura, a adição de variáveis de natureza espacial, como por exemplo a distância a algum pólo de atração, tem trazido uma melhora significativa em termos de avaliação e se apresenta como um assunto em ascensão. Porém, se no tipo de avaliação expedita utilizada por negociadores de imóveis, este tipo de variável é, de alguma forma, levado em consideração, uma modelagem matemática de sua influência ainda não é tão fácil de ser obtida.

Como pôde ser observado na revisão bibliográfica apresentada, alguns modelos de avaliação que utilizaram somente variáveis de natureza espacial não foram muito bem sucedidos. Por outro lado, observou-se alguns casos em que somente características de natureza física foram consideradas e o resultado obtido foi julgado satisfatório. Porém, também observou-se casos em que a inclusão apenas de atributos físicos não conduziu a estimativas adequadas. Diante de quadros deste tipo, fartos na literatura, a utilização de um banco de dados misto, onde tanto variáveis de natureza espacial quanto física possam ser incluídas, se torna bastante atraente.

Porém, se a inclusão de alguma medida de natureza espacial parece ser interessante, o julgamento correto de sua influência está vinculado a um método

de estimação eficiente, baseado em modelagem matemática adequada. Desta forma, além da adição de uma variável de natureza espacial, a distância ao centro da cidade, este trabalho comparou dois métodos de avaliação. O primeiro, que além de ser muito utilizado na prática, é recomendado pela ABNT (Associação Brasileira de Normas Técnicas), foi o modelo de regressão múltipla, cujos resultados serviram para comparação com o outro método utilizado aqui: as Redes Neurais. Embora casos que utilizaram este tipo de modelagem sejam escassos na literatura sobre avaliação imobiliária, os bons resultados que vêm sendo obtidos em vários campos de estudo motivaram sua utilização neste trabalho.

Para estudar a adequabilidade de uma técnica inovadora no ramo de avaliação, como é o caso das Redes Neurais, e, mais do que isso, verificar a influência da acessibilidade na valorização de imóveis urbanos, este trabalho abordou dois estudos de caso, as cidades de Araçariguama e São Carlos. A primeira é uma cidade dormitório, de pequeno porte, localizada nas proximidades da capital do estado, São Paulo. A segunda, por sua vez, é uma cidade de porte médio, localizada no centro do estado e tem como principal característica ser um pólo industrial e tecnológico. A escolha destas cidades se deu, num primeiro momento, pela facilidade de aquisição de dados, reforçada pela oportunidade de estudar a influência de uma variável como a acessibilidade em dois lugares com contextos regionais diferentes.

Os resultados obtidos para Araçariguama indicam que, mesmo sendo uma cidade de pequeno porte, a medida de acessibilidade empregada, distância ao centro, é uma das variáveis mais importantes na formação do preço de propriedades. Uma possível explicação para isso é que devido às pequenas dimensões da cidade, os pólos de atração, como escolas, bancos e área comercial se localizam no centro e as limitações do sistema de transporte público fazem com que propriedades localizadas na região central sejam mais valorizadas. Quanto aos resultados obtidos pelos modelos empregados, pôde-se observar que, ao utilizar regressão múltipla, o efeito da variável distância ao centro não pode ser estudado sozinho, pois esta variável interage com a área. Desta forma, utilizando a expressão obtida, verificou-se que um aumento na área aumenta o efeito marginal desta variável, ou seja,

propriedades grandes localizadas longe do centro são valorizadas. Uma explicação para isso é que um dos bairros mais luxuosos da cidade é composto por chácaras e se situa distante da região central.

A utilização de Redes Neurais, por sua vez, também forneceu estimativas adequadas de valor. A importância relativa de cada variável pôde ser medida através de um método proposto por Garson e forneceu um panorama da valorização de imóveis na cidade. Para os três conjuntos de dados utilizados na análise, verificou-se que a acessibilidade apresenta um peso superior a 34% no valor final, seguida pela área, com importância relativa entre 10% e 18%.

Uma comparação dos dois métodos para Araçariguama mostrou que, em termos do coeficiente de determinação (R^2), não se pode concluir muita coisa, uma vez que estes valores oscilaram muito de um conjunto para outro de dados. Ao considerar o valor da erro relativo total, no entanto, pode-se observar que as Redes Neurais (RN) demonstraram um desempenho superior, gerando uma diminuição de 43% com relação ao modelo de regressão. Uma comparação ponto a ponto para os erros obtidos pelos dois métodos e os três conjuntos de dados analisados indicou que, no geral, os erros relativos fornecidos pelas RN foram menores.

Com o objetivo de analisar a distribuição espacial dos erros fornecidos pelo modelo de regressão, estes foram agrupados em cinco *clusters*. Os mesmos limites de classe foram utilizados para agrupar os erros fornecidos pelas RN. Observou-se daí que, no geral, a maior proporção de erros observados pelo emprego das Redes Neurais se concentrou nos *clusters* com valores menores. Ao observar os mapas temáticos gerados, pode-se constatar que os maiores erros fornecidos pelas RN se concentram em um único bairro. Para a regressão, no entanto, não se consegue distinguir uma concentração espacial dos erros maiores.

A análise para São Carlos seguiu o mesmo critério utilizado para a outra cidade. Assim como para Araçariguama, pôde-se constatar que a distância ao centro é um dos fatores preponderantes na avaliação imobiliária local, o que pôde ser observado pelos resultados obtidos por ambos os métodos. Os resultados obtidos indicaram que um aumento no valor da distância ao centro da cidade provoca uma diminuição do preço por metro quadrado.

Uma análise da importância de cada variável na resposta estimada para as RN mostrou que, também para São Carlos, a acessibilidade é a variável que contribuiu com a maior representatividade na estimação de valor em dois dos conjuntos de dados, seguida pela categoria de bairro a que pertence o lote. Diferentemente do que aconteceu para Araçariçuama, não se percebeu uma diferença significativa entre os valores estimados por este método e os fornecidos pela regressão, tanto em termos de coeficiente de determinação quanto de erro relativo total.

Através dos mapas temáticos construídos para verificar a distribuição espacial dos erros, notou-se uma concentração de erros maiores em um dos bairros analisados, o que pôde ser observado para os dois métodos empregados e dois dos conjuntos de dados. Uma possível explicação para este problema é que o bairro em questão possui uma grande quantidade de lotes vazios sendo oferecidos a preços de mercado bastante baixos quando comparados aos outros bairros da cidade. Uma solução para este problema poderia ser a divisão da décima quarta variável (categoria do bairro) em mais um subgrupo, indicando se o terreno se localiza neste bairro. Além dessa, outra opção seria isolar estes lotes do conjunto de dados e construir modelos diferentes para eles. Tais sugestões, que merecem uma atenção especial, sugerem um estudo mais aprofundado das características do bairro em questão.

Além da comparação dos resultados fornecidos por ambos os métodos, uma análise do desempenho das Redes Neurais também foi realizada. Para tal, redes utilizando o valor total do lote em lugar do preço por m^2 , foram simuladas para as duas cidades. Os resultados indicaram que, se o objetivo for apenas o valor estimado e não a influência de cada variável, uma configuração de rede que conduz aos menores erros relativos para uma situação, também pode ser utilizada para a outra. Porém, os erros relativos obtidos quando o preço total do lote foi utilizado apresentaram, no geral, valores maiores que os obtidos para o modelo que utilizou o m^2 .

Além da diferença observada nos valores dos erros, notou-se que a área do lote passou a ter a maior representatividade nos valores estimados para o

preço total, fato observado para os dois estudos de caso e que já era esperado devido ao tipo de variável utilizada. Além disso, os mapas temáticos construídos para a verificação da distribuição espacial dos erros fornecidos para o caso em que o valor total do lote foi utilizado para a cidade de São Carlos, apresenta uma certa coerência com aqueles obtidos para os erros observados para os valores estimados do m^2 , sugerindo, mais uma vez que, em termos de valor estimado e para este estudo, as redes forneceram boas estimativas, independentemente da transformação a que as variáveis foram previamente submetidas.

Quanto à acessibilidade, um fato que chamou a atenção foi o de que para Araçariguama, que é uma cidade de porte menor, a importância relativa desta variável foi maior que para a outra cidade. Uma possível explicação para isso é que, além da existência de um sistema de transporte público com boa cobertura, São Carlos apresenta alguns pólos de atração em outros bairros, o que faz com que moradores destes lugares não precisem se deslocar até a região central da cidade para executar várias tarefas e, portanto, a variável distância ao centro parece não ser tão decisiva na aquisição de uma propriedade. Este fato sugere que outros tipos de medida de acessibilidade, que levem em consideração situações deste tipo, também devem ser analisados. A comparação de cidades com portes e características semelhantes também é um assunto que deve ser futuramente considerado, com o objetivo de verificar se a influência da acessibilidade ocorre de maneira similar.

A utilização de uma ferramenta mais poderosa de estatística espacial também mostrou-se de grande importância para a análise de resultados, uma vez que poderia executar o trabalho de visualização da distribuição dos erros de maneira mais adequada. Apesar da simplicidade do método aqui utilizado, que uniu ferramentas estatísticas a um Sistema de Informações Geográficas, pôde-se observar fatos que haviam passado despercebidos, como por exemplo, a concentração de erros maiores em determinados bairros. Problemas como este poderiam ser facilmente identificados e quantificados se fossem empregadas ferramentas específicas para este tipo de análise.

De maneira geral, os resultados obtidos aqui sugerem que, em termos de facilidade de aplicação, os programas que fazem uso de Redes Neurais se

apresentam como um método mais acessível, em termos de interface com o usuário. Por sua vez, a verificação de algumas suposições básicas com relação aos modelos de regressão, embora raramente executada pelos usuários deste tipo de método, restringe seu uso a pessoas que tenham noções de álgebra de matrizes e nem sempre estão disponíveis em pacotes computacionais específicos, o que dificulta sua utilização de forma teoricamente correta.

O inconveniente das RN, por outro lado, é a obscuridade da lógica interna de processamento, em geral não visível para os usuários. Isto é, de certa forma, desmistificado pelo método de Garson aqui empregado.

Diante dos resultados comentados aqui, sugere-se a seguinte relação de trabalhos para futuro desenvolvimento:

- Avaliação do emprego de outras medidas de acessibilidade;
- Comparação da influência da acessibilidade para cidade de portes e características parecidas;
- Inclusão de uma nova categoria de bairro no banco de dados de São Carlos e novo treinamento da rede;
- Utilização de ferramentas específicas de análise espacial para a verificação da distribuição dos erros.

7. BIBLIOGRAFIA

- ABELSON, P. (1997). *House and Land Prices in Sydney from 1931 to 1989*. In: Urban Studies, v. 34, n. 9. pp. 1381-1400.
- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 5676/90. (1989). *Avaliação de Imóveis Urbanos*. Rio de Janeiro, ABNT.
- ALMOND, N.; JENKINS, D.; GRONOW, S. (1997). *Development of a Prototype Residential Valuation System*. In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 1. pp. 204-215.
- ANTON, H. (1991). *Elementary Linear Algebra*. Segunda edição. John Wiley & Sons. 526 p.
- AOKI, Y.; OSARAGI, T.; ISHIZAKA, K. (1994). *An Interpolating Model for Land-price Data with Transportation Costs and Urban Activities*. In: Environment and Planning, v. 21. pp. 53-65.
- AZAR, K. T.; FERREIRA JR, J.; ABED J.; KAZZAZ, T. (1994). *Using GIS for Estimating the Spatial Distribution of Land Value in Metropolitan Beirut*. Urban and Regional Information Association - URISA.
- BALARINE, O. F. O. (1995). *Contribuições Macroeconômicas ao Entendimento da Formação de Preços Habitacionais Locais*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis. Anais... Florianópolis, 1995. pp. 361-265.
- BEALE, R.; JACKSON, T. (1990). *Neural Computing: An Introduction*. Institute of Physics Publishing.
- BENTICK, B. L. (1996). *The Differential Incidence of an Urban Land Tax Depends on the Travel Intensities of Substitutes for Land*. In: Urban Studies, v. 33, n. 9. pp. 1729-1732.
- BIELLI, M.; REVERBERI, P. (1996). *New Operations Research and Artificial Intelligence Approaches to Traffic Engineering Problems*. In: European Journal of Operational Research, n. 92, pp. 550-572.
- BORGES, J. M. M. (1975). *Custo de Transporte e Valor dos Terrenos Urbanos*. Dissertação de Mestrado. Faculdade de Economia e Administração - Universidade de São Paulo.

- BRANDÃO, D. Q.; GUCH, D. U.; NEUMANN, C. (1995). *Avaliação de Unidades de Edifícios Residenciais Utilizando Estatística Inferencial: Aplicação em Bairros da Região Continental de Florianópolis*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis. Anais... Florianópolis, 1995. pp. 229-236.
- BRONDINO, N. C. M.; SILVA, A. N. R. (1997). *Comparando Estratégias de Avaliação de Imóveis Para Fins de Tributação Com o Auxílio de um SIG*. GIS Brasil 1997 Curitiba. Anais... Curitiba, 1997. (CD-ROM).
- BRONDINO, N. C. M.; SILVA, A. N. R. (1998a). *Atualização de Plantas Genéricas de Valores: Novo Levantamento Cadastral ou Amostra de Valores de Mercado?* GIS Brasil 1998. Curitiba. Anais... GIS Brasil 98. (CD-ROM).
- BRONDINO, N. C. M.; SILVA, A. N. R. (1998b). *A Comparison of Land Valuation Methods Supported by GIS*. In: 4th Conference on Design and Decision Support Systems in Architecture and Urban Planning. Maastricht. Anais... Maastricht, 1998 (CD-ROM).
- BULLEN, N. (1997). *Linking GIS and Multilevel Modelling in the Spatial Analysis of House Prices*. In: CRAGLIA, M. & COUCLELIS, H. *Geographical Information Research*. pp. 412-426.
- BUZAI, G. D.; DURAN, D. (1997). *Enseñar e Investigar com Sistemas de Informação Geográfica (S.I.G.)*. Editora Troquel. 191 p.
- CÂMARA, G.; CASANOVA M. A.; HEMERLY, A. S.; MAGALHÃES, G. C.; MEDEIROS, C. M. B. (1996). *Anatomia de Sistemas de Informação Geográfica*. Instituto de Computação, UNICAMP. Campinas. 1996. 193p.
- CARVALHO, A. P. L. F.; BRAGA, A.; LUDERMIR, T. (1998). *Fundamentos de Redes Neurais*. Publicação da COPPE/UFRJ para a XI Escola de Computação Brasileira. Rio de Janeiro.
- CASSIDY, H. J. (1981). *Using Econometrics: A Beginner's Guide*. Reston Publishing Company. Virgínia. 304p.
- CHESHIRE, P.; SHEPPARD, S. (1995). *On the Price of Land and the Value of Amenities*. In: *Economica*. v. 62. pp. 247-267.
- COCHRAN, W. G. (1953). *Sampling Techniques*. John Wiley & Sons. pp. 65-90.
- DANTAS, R. A. (1999). *Engenharia de Avaliações – Uma Introdução à Metodologia Científica*. Editora Pini. São Paulo. 251p.
- DAVIDSON, K. B. (1995). *Accessibility and Isolation in Transport Network Evaluation*. In: WCTR –The Seventh World Conference on Transport Research. Sydney. pp. 10-8.
- DIA, H.; ROSE, G. (1995). *Development of Artificial Neural Network Models for Automated Detection of Freeway Incidents*. In: WCTR –The Seventh World Conference on Transport Research. Sydney. pp. 1-21.
- DRAPER, N. R.; SMITH, H. (1980). *Applied Regression Analysis*. John Wiley & Sons. 709 p.
- DUNN, O. J.; CLARK, V. A. (1987). *Applied Statistics: Analysis of Variance and Regression*. Second Edition. John Wiley & Sons. 445 p.

- DUARTE, A. M.; GABBAY, A. (1995). *Avaliação em Massa: “Modelos Genéricos de Valores Integrados a um Sistema de Informações Geográficas - SIG”*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis. Anais... Florianópolis, 1995. pp. 237-240.
- FALAS, T. (1995). *Neural Networks in Empirical Accounting Research: An Alternative to Statistical Models*. In: Neural Network World, v. 5, n. 4. pp. 419-432.
- FERRAZ, A. C. P. (1990). *Por um Imposto Sobre os Terrenos Urbanos Ociosos*. Folha de São Paulo. São Paulo, 26 de jan. cad. B, p.2.
- FERRAZ, A. C. P.; SILVA, A. N. R. (1990). *Reforma Urbana Já*. Folha de São Paulo. São Paulo, 01 de mai. cad. C, p.5.
- FERRAZ, A. C. P.; SILVA, A. N. R. (1992). *Tributação Justa de Propriedade*. Folha de São Paulo. São Paulo, 24 de set. cad. 3, p.5.
- FERRARI, R. (1996). *Cenários Alternativos para o Uso de GIS em Municípios*. GIS BRASIL 96. Curitiba. Anais ... Curitiba, 1996. Módulo Municipal. pp.17-23.
- FIKER, J. (1989). *Manual de Redação de Laudos*. Editora Pini. São Paulo. 119 p.
- FIKER, J. (1993). *Avaliação de Imóveis Urbanos*. Quinta Edição. Editora Pini. São Paulo. 104 p.
- FLITMAN, A. M. (1997). *Towards Analyzing Student Failures: Neural Networks Compared with Regression Analysis and Multiple Discriminant Analysis*. In: Computers & Operations Research, v. 24, n. 4. pp. 367-377.
- FORREST, D.; GLEN, J.; WARD, R. (1996). *The Impact of a Light Rail System on the Structure of House Prices – A Hedonic Longitudinal Study*. In: Journal of Transport Economics & Policy, v. 30, n.1. pp. 15-29.
- GIBB, K. (1994). *Before and After Deregulation: Market Renting in Glasgow and Edinburgh*. In: Urban Studies, v. 31, n. 9 pp. 1481-1495.
- GONÇALVES, M. F. R.. (1988). *O Imposto Predial e Territorial Urbano e a Progressividade*. In: Revista de Administração Municipal. Rio de Janeiro, v. 35, n.189. pp. 6-14, out./dez.
- GONÇALVES, M. F. dos R. (1990). *IPTU: Alcance dos Princípios da Legalidade e da Anualidade*. In: Revista de Administração Municipal. Rio de Janeiro, v. 37, n. 197. pp. 54-61, out./dez.
- GONZÁLEZ, M. A. S. (1995) *Plantas de Valores Inferenciais: A Espacialidade Considerada Através de Trend Surfaces*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis. Anais... Florianópolis, 1995. pp. 390-397.
- GONZÁLEZ, M. A. S. (1996a). *Aspectos Legais e Conceituais da Fixação de Valores do IPTU*. In: II COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais... Florianópolis 1996. pp. 123-130.
- GONZÁLEZ, M. A. S. (1996b) *Desenvolvimento de Planta de Valores com Sistemas de Informações Geográficas e ITBI*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996. Tomo II. pp. 225-231.

- GONZÁLEZ, M. A. S.; FORMOSO, C. T. (1994) *O Emprego de Dados do ITBI Para Atualização de Cadastros de Valores Venais: Um Estudo de Viabilidade*. In: I COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário - Florianópolis, 1994. Anais... Florianópolis. Módulo CTMU. pp. 154-162.
- GONZÁLEZ, M. A. S.; FORMOSO, C. T. (1995). *Integração dos Tributos Imobiliários Urbanos: Uma Planta de Valores Inferencial Única, Determinada com Dados do ITBI*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis, 1995. Anais... Florianópolis, 1995, p. 326-335.
- GONZÁLEZ, M. A. S. (1993). *A Formação do Valor de Aluguéis de Apartamentos Residenciais na Cidade de Porto Alegre*. Dissertação de Mestrado. Porto Alegre. CPGEC/UFRGS.
- GREENE, W. H. (1997). *Econometric Analysis*. Third Edition. Prentice Hall. New Jersey. 1075 p.
- GUEDES, J. C. (1995). *O Emprego de Inteligência Artificial na Avaliação de Bens*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis. Anais... Florianópolis, 1995. pp. 368-374.
- GUNST, R. F.; MASON, R. L. (1980). *Regression Analysis and its Application – A Data Oriented Approach*. Marcel Dekker. New York. 402 p.
- HADDAD, E. (1995). *Using GIS to Monitor Housing Markets in São Paulo, Brazil*. In: 4th International Conference on computers in Urban Planning and Urban Management. Melbourne. Anais...Melbourne 1995, v.2. pp. 481-494.
- HADDAD, E. (1997). *Using GIS to Monitor Housing Markets in São Paulo, Brazil: Part II*. In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 1 pp. 193-203.
- HAINING, R. (1995). *Designing Spatial Data Analysis Modules for Geographical Information Systems*. In: FOTHERRINGHAN, S. and ROGERSON, P. *Spatial Analysis and GIS*. pp. 45-63.
- HARDGRAVE, B. C.; WILSON, R. L.; WALSTROM, K. A. (1994). *Predicting Graduate Students Success: A Comparison of Neural Networks and Traditional Techniques*. In: Computers & Operations Research, v. 21, n. 3. pp. 249-263.
- HAYKIN, S. (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan College Publishing Company. pp. 1-88.
- HOCHHEIM, N.; PETRY, G.; SCHERER, S. R.; SIMON, L. M.; WOLSKI, M. S. (1996) *Avaliação de Imóveis Urbanos por Estatística Inferencial e Regressão Múltipla: Aplicação para o Centro de Florianópolis*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996. Tomo II pp. 61-69.
- HSIAO, S.; STERLING, J. (1992). *Use of Geographic Information System for Transportation Data Analysis*. In: Chow, J.; Litvin, D. M.; Opiela, K. S. (eds). Microcomputers in Transportation. American Society of Civil Engineers. New York. pp 94-102.
- HUXHOLD W. E. (1991). *An Introduction to Urban Geographic Information Systems*. Oxford University Press. Wisconsin. 337p.

- JOHNSON, R. A.; WICHERN, D. W. (1998). *Applied Multivariate Statistical Analysis*. Fourth Edition. Prentice Hall. 816 p.
- KIRCHNER, F. F. (1994). *O Sistema Geográfico de Informações e o Cadastro Técnico Multifinalitário*. In: I COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais... Florianópolis, 1994.
- KWON, O.; GOLDEN, B.; WASIL, E. (1995). *Estimating the Length of the Optimal TSP Tour: An Empirical Study Using Regression and Neural Networks*. Computers & Operations Research, v. 22, n. 10. pp. 1047-1056.
- LAPOLLI, A. R. S.; DE CESARE, C. M.; LUNARDI, M. L. F.; OLIVEIRA, O. S. de; GRANDO, P. A. (1994). *Metodologia para a Determinação de Regiões Homogêneas de Valorização Imobiliária, Tendo em Vista a Geração de Informações Cadastrais: O Caso do Município de Porto Alegre*. In: I COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais... Florianópolis, 1994, Módulo CTMU.
- LEVY, P. S.; LEMESHOW, S. (1991). *Sampling of Populations-Methods and Applications*. John Wiley & Sons.
- LEVIN, E. J.; WRIGHT, R. E. (1997). *Speculation in the Housing Market?* In: Urban Studies, v. 34, n.9. pp. 1419-1437.
- LEVINE, N. (1996). *Spatial Statistics and GIS – Software Tools to Quantify Spatial Patterns*. In: Journal of The American Planning Association, v. 62, n. 3. pp. 381-391.
- LIM, D. (1992). *The Nonneutrality of the Land Value Tax: Impacts on Urban Structure*. In: Journal of Urban Economics, v. 32. pp. 186-194.
- LIPSCHULTZ, S. (1968). *Linear Algebra*. Schaum Publish Co. 413 p.
- LONGLEY, P.; BATTY, M. (1996). *Spatial Analysis: Modelling in a GIS Environment*. John Wiley & Sons. 392 p.
- LONGO, C. A. (s.d.). *Finanças Municipais e Ociosidade dos Terrenos na Área Urbana*. In: LONGO, C. A. & RIZZIERI, J. A. B. org. *Economia urbana: custos de urbanização e finanças públicas*. Fundação Instituto de Pesquisas Econômicas - Série Relatórios de Pesquisa RP-10. São Paulo. pp. 197-232.
- LONGO, C. A.; LIMA, J. C. S. (s.d.). *O IPTU como Fonte de Recursos a Nível Municipal: Aspectos de Eficiência e Equidade*. In: LONGO, C. A. & RIZZIERI, J. A. B., org. *Economia urbana: custos de urbanização e finanças públicas*. Fundação Instituto de Pesquisas Econômicas - Série Relatórios de Pesquisa RP-10. São Paulo. pp. 233-256.
- LUGER, G. F.; STUBBLEFIELD, W. A. (1989). *Artificial Intelligence and the Design of Expert Systems*. The Benjamin/Cummings Publishing Company. pp. 581-584.
- MEDEIROS JR., J. R.; FIKER, J. (1996). *A Perícia Judicial – Como Redigir Laudos e Argumentar Dialecticamente*. Editora Pini. São Paulo. 138 p.
- MENDONÇA, M. C.; SOLLERO FILHO, M. ; CURI, E.; AGUIAR, J. B.; QUEIROGA, H. S.; MAIA, E. A.; AQUINO, R.; RESENDE, O.; CANÇADO, J. M. M. (1998). *Fundamentos de Avaliações Patrimoniais e Perícias de Engenharia*. Editora Pini. São Paulo. 316 p.
- MÖLLER, L. F. C. (1995). *Planta de Valores Genéricos – Avaliação de Imóveis Para Fins Tributários*. Sagra – DC Luzzatto Editores. Porto Alegre. 79 p.

- MOREIRA, A. L. (1997). *Princípios de Engenharia de Avaliações*. Quarta Edição. Editora Pini. São Paulo. 504 p.
- MUSSONI, L.; RINELLI, S. REITANI, G. (1996). *Estimating the Accident Probability of a Vehicular Flow by Means of an Artificial Neural Network*. In: Environment and Planning B: Planning and Design, v 23, n. 6. pp. 667-676.
- NATH, R.; RAJAGOLAPALAN, B.; RYKER, R. (1997). *Determining the saliency of input variables in Neural Network Classifiers*. In: Computers & Operational Research, v. 24, n. 8. pp. 767-773.
- OTTENSMANN, J. R. (1977). *Urban Sprawl, Land Values and the Density of Development*. In: Land Economics, v. 53, n. 4. pp. 389-400.
- PAULINO, A. A. D. D.; FREITAS, A. F; HEINECK, L. F. M. (1995) *Atributos Valorativos dos Imóveis e Pesquisas Mercadológicas: Uma Análise de Mercado de Florianópolis - SC*. In: VIII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias. Florianópolis. Anais... Florianópolis, 1995. pp. 176-184.
- PERUZZO, C. A.; LIMA JR., C. O.; RECH, J. V.; MARISCO, N. (1996) *Avaliação de Apartamentos por Inferência Estatística - Estudo de Caso: Balneário de Canasvieiras, Florianópolis - SC*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996. Tomo II pp. 152-158.
- PHANG, S.; WONG, W. (1997). *Government Policies and Private Housing prices in Singapore*. In: Urban Studies, v. 34, n. 11. pp. 1819-1829.
- PASHA, H. A.; BUTT, M. S. (1996) *Demand for Housing Attributes in Developing Countries: A Study of Pakistan*. In: Urban Studies, v. 33, n. 7. pp. 1141-1154.
- PATKAR, V. N. (1997) *On Integration of Analytical Techniques and GIS for Urban Management*. In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 1. pp.257-268.
- PRATT, A. C. (1996). *Coordinating Employment, Transport and Housing in Cities: An Institucional Perspective*. In: Urban Studies, v. 33, n. 8. pp. 1357-1375.
- RAIA JR, A. A. (1995). *Uma Avaliação do Modelo META para Cálculo de Custos de Transportes e Seu Uso na Tributação de Terrenos Urbanos Ociosos..* Dissertação de Mestrado - Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos. 144 p.
- RAIA JR, A. A.; SILVA, A. N.R.; LIMA, R. S. (1996a). *Acessibilidade e Valor de Terrenos Urbanos - O Caso da Cidade de Araraquara*. In: VIII CLATPU - Congresso Latino Americano de Transporte Público e Urbano. Curitiba. Anais... Curitiba 1996. pp. VI-1 - VI-11.
- RAIA JR., A. A.; SILVA, A. N. R.; LIMA, R. S. (1996b). *Utilizando um SIG Para Avaliar Níveis de Acessibilidade em Uma Cidade Média*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996, Tomo III. pp. 193-204.
- RAIA JR., A. A.; SILVA, A. N. R. (1996). *Relação Acessibilidade-Preço de Terrenos Com o Uso de um SIG*. In: CONGRESSO TÉCNICO-CIENTÍFICO DE ENGENHARIA CIVIL. Florianópolis. Anais... Florianópolis, 1996. v. III. pp.338-348.

- RAIA JR. A. A. (1997). *Acessibilidade e Preço do Terrenos em Cidades Médias de um País em Desenvolvimento*. Exame de qualificação ao doutorado apresentado ao Departamento de Transportes da EESC – USP.
- RIPLEY, B. D. (1981). *Spatial Statistics*. John Wiley & Sons. New York. 252 p.
- ROAKES, S. L. (1996). *Reconsidering Land Value Taxation – The Golden Key?* In: Land Use Policy, v. 13, n. 4. pp. 261-272.
- RODGHER, S. F.; FABBRI, G. T. P.; CARVALHO, A. C. P. L. F. (1997). *A Utilização de Redes Neurais Artificiais para Classificação de Solos Tropicais*. In: XI Congresso de Pesquisa e Ensino em Transportes. Rio de Janeiro.
- SALCHENBERGER, L.; VENTA, E. R.; VENTA, L. A. (1997). *Using Neural Networks to Aid the Diagnosis of Best Implant Rupture*. In: Computers & Operations Research, v. 24, n. 5. pp. 435-444.
- SAMPAIO, A. C. F. (1996) *Mapeamento Cadastral - Exemplos e Alternativas de Execução*. In: II COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais... Florianópolis, Tomo III. pp. 386-392.
- SCHNEIDER, V. P. (1994). *Cadastro Técnico Multifinalitário e o Sistema Tributário*. In: I COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário - Florianópolis. Anais... Florianópolis 1994.
- SILVA, A. N. R. (1993). *O Custo do Solo Urbano Ocioso e uma Nova Sistemática de Tributação da Propriedade*. Tese de Doutorado em Transportes, Universidade de São Paulo. São Carlos. 137 p.
- SILVA, A. N. R. (1998). *Sistemas de Informação Geográfica em Transportes*. Texto apresentado à Escola de Engenharia de São Carlos para obtenção do título de Livre-Docente. São Carlos. 112 p.
- SILVA, A. N. R.; FERRAZ, A. C. P. (1993). *Uma Nova Sistemática de Tributação da Propriedade Urbana*. In: Revista de Administração Municipal. IBAM - Instituto Brasileiro de Administração Municipal, v. 40, n. 208. pp. 51-61, jul/set.
- SILVA, A. N. R.; VARGAS, E. N. Z.; ITO, L. E.; CUBAS, S. I. B. (1994). *O Cadastro dos Terrenos Ociosos e a sua Importância para uma Tributação Mais Justa da Propriedade Urbana*. In: I CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO. Florianópolis. Anais... Florianópolis, 1994, Tomo III pp.118-126.
- SILVA, A. N. R.; MELO, J. J. O.; BRONDINO, N. C. M. (1997). *Uma Introdução ao Planejamento de Transportes com Sistemas de Informação Geográfica*. Editora da Escola de Engenharia de São Carlos. São Carlos. 74p.
- SILVA, A. N. R.; WAERDEN, P. v. (1997) *Combining Simplified Models and GIS to Estimate Urban Costs in Developing Countries*. In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 1. pp. 85-96.
- SILVA, A. N. R.; FERRAZ, A. C. P.; RAIA JR., A. A. (1999). *Towards a Strategy to Minimize the Negative Effects of Urban Sprawl in Brazil*. In: Jenks, M. & Burgess, R. The Compact City in Developing Countries. Routledge. Londres. (no prelo).

- SUBRAMANIAN, V.; HUNG, M. S.; HU, M. Y. (1993). *An Experimental Evaluation of Neural Networks for Classification*. In: Computers & Operations Research, v. 20, n. 7. pp. 769-782.
- TEIXEIRA, A. L. A.; MORETTI, E.; CHRISTOFOLETTI, A. (1992). *Introdução aos Sistemas de Informação Geográfica*. Edição do autor. Rio Claro.
- TOLEDO, M. C. S.; DOMINGUES, R. M. (1995). *Valores Unitários Básicos para Glebas e Lotes no Estado de São Paulo*, VII COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias. Florianópolis. Anais... Florianópolis, 1995. pp. 448-456.
- TRIPPI, R. R.; TURBAN, E. (1992). *Neural Networks in Finance and Investing*. Probus Publishing Company. pp. 3-72.
- TRIVELLONI, C. A.; LIMA JUNIOR, C. O.; RECH, J. V.; MARISCO, N. (1996). *Avaliação de apartamentos por inferência estatística - estudo de caso: Balneário de Canasvieiras, Florianópolis - SC*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996. Tomo II. pp. 152-160.
- TUBB, N. R. (1993). *A Development Path to Success in Neural Computing*. In: Expert Systems Applications, v. 9, n. 5. pp. 5-9.
- ULUACHA, N. O. (1997). *The Development and Application of a Geographical Information system (GIS) for Property Tax Administration: Experiences from Nigeria* In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 2. pp. 560-571.
- VERTELO, J. B. (1996). *Planta Genérica de Valores da Cidade de Governador Valadares*. In: II COBRAC - Cadastro Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais.... Florianópolis, 1996, Tomo II. pp. 232-243.
- WONG, D. W. S.; WU, V. C. (1995). *Quality of Aggregated Spatial Data – A Guidance for Decision*. In: 4th International Conference on Computers in Urban Planning and Urban Management. Melbourne. Anais...Melbourne 1995, v.2. pp. 559-570.
- WYATT, M. D. (1994). *A Critical View of Land Value Taxation as a Progressive Strategy for Urban Revitalization, Rational Land use, and Tax Relief*. In: Review of Radical Political Economics, v. 26, n. 1. pp. 1-25.
- WYATT, R. (1995). *Strategy Evaluation Using Neural Networks*. In: Proceedings of the 4th International Conference in Urban Planning and Urban Management. Melbourne, pp. 367-378.
- WYATT, R. (1996). *Evaluating Strategies by Means of an Artificial Neural Network*. In: Environment and Planning B: Planning and Design, v. 23, n. 6. pp. 685-696.
- WYATT, R. (1997). *Consistency of Planning Style*. In: Proceedings of the 5th Conference on Computers in Urban Planning and Urban Management, Mumbai. Anais... Mumbai, 1997, v. 1, pp.181-192.
- XU, F.; MITTELHAMMER, R. C.; BARKLEY, P. W. (1993). *Measuring the Contributions of Site Characteristics to the Value of Agricultural Land*. In: Land Economics, v. 69, n. 4. pp. 356-369.
- YUAÇA, F. (1996). *Estimador de Valor do Lote Urbano*. GIS BRASIL 96. Curitiba. Anais ... Curitiba, 1996. Módulo Municipal. pp.25-39.

ZANCAN, E. C.; HEINECK, L. F. (1994). *Metodologia Para Execução de Planta de Valores Genéricos: Caso de Criciúma-SC*. In: I COBRAC - Congresso Brasileiro de Cadastro Técnico Multifinalitário. Florianópolis. Anais... Florianópolis, 1994. Módulo CTMU. pp. 52-59

ZANCAN, E. C. (1996). *Avaliações de Imóveis em Massa Para Efeitos de Tributos Municipais*. Rocha Gráfica e Editora Ltda. Florianópolis. 121p.