

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
Área de concentração: Engenharia de Computação e Telecomunicações
Linha de pesquisa: Inteligência Computacional

**REDES NEURAS ARTIFICIAIS
APLICADAS ÀS AVALIAÇÕES EM MASSA
ESTUDO DE CASO PARA A CIDADE DE
BELO HORIZONTE / MG**

Belo Horizonte / MG

Março 2006

Antônio Pelli Neto

**REDES NEURAIS ARTIFICIAIS
APLICADAS ÀS AVALIAÇÕES EM MASSA
ESTUDO DE CASO PARA A CIDADE DE
BELO HORIZONTE / MG**

Dissertação apresentada ao programa de pós-graduação da Universidade Federal de Minas Gerais, como requisito para a obtenção de título de Mestre em Engenharia Elétrica, sob a orientação do Prof. Antônio de Pádua Braga, PhD.

Área de concentração: Engenharia de Computação e Telecomunicações.

Linha de Pesquisa: Inteligência Computacional.

Belo Horizonte / MG

Março 2006

Pelli Neto, Antônio.

Redes Neurais Artificiais aplicadas às Avaliações em Massa - Estudo de caso para a cidade de Belo Horizonte / MG.

Dissertação de Mestrado – Universidade Federal de Minas Gerais – UFMG.

1. Mercado Imobiliário
2. Estimação de Preços Hedônicos
3. Econometria Espacial
4. Redes Neurais Artificiais
5. Estudo de Caso – Belo Horizonte/ MG.

*Esta dissertação é dedicada à minha família:
Dolores, André, Júlia e Renato.*

“Existe um tempo para melhorar, para se preparar e planejar; igualmente existe um tempo para partir para a ação...”.

“Porque um dia é preciso parar de sonhar, tirar os planos das gavetas e, de algum modo, começar”.

Amyr Klink

Agradecimentos

São Tomás de Aquino ensina que a gratidão é uma realidade humana complexa: A gratidão se compõe de diversos graus. “O primeiro consiste em reconhecer o benefício recebido; o segundo, em louvar e dar graças; o terceiro, em retribuir de acordo com suas possibilidades e segundo as circunstâncias mais oportunas de tempo e lugar”.

Ao finalizar o curso de mestrado, não posso deixar de reconhecer todo o benefício recebido. E estes benefícios somente puderam ser obtidos graças ao apoio de diversas pessoas que, com certeza, não conseguirei aqui relacioná-las por completo.

Desejo inicialmente agradecer aos meus pais, que sempre acreditaram que a educação e os estudos são fundamentais em nossas vidas. E especialmente à minha mãe, pela saudade deixada com sua partida desta vida.

À minha família, com carinho especial à minha esposa Dolores, pelo seu inesgotável esforço em me apoiar, assumindo muitas vezes atividades que eu deveria estar compartilhando. Sua determinação, competência e a atenção dedicada durante estes dois anos são dádivas por mim recebidas, e que tenho um profundo agradecimento. Aos meus filhos, André, Júlia e Renato, deixo aqui o reconhecimento de que estive ausente em momentos importantes em suas vidas, mas que também assumo o compromisso de resgatá-los.

Ao meu orientador, neste mestrado, o Prof. Antonio de Pádua Braga, principalmente pelo apoio dado nos momentos mais difíceis, se mostrando, além de Doutor em sua área de atuação, um grande motivador.

Aos colegas da Caixa Econômica Federal, Lilia, Dimas, Ernane e Jacqueline, enfim, a todos os que direta ou indiretamente me apoiaram.

À Caixa Econômica Federal, pelo fornecimento das informações sobre o mercado imobiliário e à Prefeitura Municipal de Belo Horizonte, pelas referências espaciais dos dados amostrados.

Por fim, é necessário agradecer à Universidade Federal de Minas Gerais, ao seu corpo docente e aos demais colegas, que me propiciaram um ambiente de ensino de alta qualidade. À sociedade brasileira, desejo em breve retribuir as oportunidades oferecidas.

Finalizo com o agradecimento ao nosso Criador, que sempre ilumina os nossos caminhos.

Redes Neurais Artificiais
aplicadas às avaliações em massa: estudo de caso para a cidade
de Belo Horizonte / MG

Este trabalho tem o objetivo de mostrar a importância da aplicação das Redes Neurais Artificiais – RNAs nos estudos dos fenômenos mercadológicos, e em particular ao que se aplica na Engenharia de Avaliações para a interpretação do mercado imobiliário. A superioridade destes modelos em relação aos estimados pela econometria tradicional foi comprovada.

A construção de modelos de formação de preços imobiliários por meio das técnicas convencionais, como é atualmente realizada, enfrenta problemas que diminuem a precisão das estimativas de valores, especialmente pelo desconhecimento da forma funcional que descreve a relação entre as variáveis e pela dificuldade da estimação dos parâmetros relativos à distribuição espacial dos imóveis. A aplicação das redes neurais artificiais em avaliação de imóveis apresenta boas perspectivas e os resultados obtidos até o momento indicam que as RNAs são ferramentas computacionais melhores para áreas onde os dados não podem ser representados unicamente por modelos lineares.

Como a determinação do valor de imóveis é empregada em um grande número de situações, inclusive para tributação, os eventuais erros cometidos podem afetar expressiva parcela da população. Assim, neste trabalho conclui-se pela importância da pesquisa de técnicas alternativas, no caso as RNAs, às atualmente utilizadas, de modo a aprimorar o processo de avaliação de bens.

Summary of a Dissertation presented at CPDEE / UFMG.

Artificial Neural Networks applied to mass evaluation: case study for Belo Horizonte city / MG

The purpose of this paper is to show the importance of the Artificial Neural Networks – ANN application on the study of the marketing phenomena, and particularly concerning its application on the real estate market interpretation in the Evaluation Engineering. These models superiority regarding the estimations by the traditional econometrics has already been proved.

The construction of real estate pricing models through conventional techniques as they are currently performed faces problems that reduce the estimation precision, particularly due to the ignorance about the functional form that describes the relation between the variables and due to the difficulty concerning the parameters estimation related to the real estate space distribution. Artificial neural networks application in real estate evaluation offers good perspectives and the outcomes so far have shown that ANN are better computational tools than regression linear for the areas where data cannot be represented solely by linear models.

As the determination of real estate value is used in a great number of situations, including for taxation, the occasional mistakes may affect expressive parcel of the population. Therefore, this paper concludes for the importance of researches on the alternative techniques use, such as ANN, rather than those currently in use, so as to improve the real estate evaluation process.

Sumário

CAPITULO 1	1
1. INTRODUÇÃO.....	1
1.1 PRELIMINARES	1
1.2 OBJETIVOS DO TRABALHO	5
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO.....	5
CAPITULO 2	7
2. MERCADO IMOBILIÁRIO – CONCEITOS E REVISÃO LITERÁRIA	7
2.1 INTRODUÇÃO	7
2.2 HISTÓRICO DA AVALIAÇÃO.....	8
2.3 LEIS E REGULAMENTAÇÕES.....	11
2.4 MERCADO IMOBILIÁRIO	12
2.5 VALOR DE MERCADO, PREÇO E CUSTO	14
2.6 ENFOQUES BÁSICOS NA AVALIAÇÃO DE BENS.....	15
2.7 METODOLOGIAS APLICÁVEIS.....	15
2.8 O MÉTODO COMPARATIVO DE DADOS DE MERCADO.....	16
2.9 CONSTRUÇÃO DAS VARIÁVEIS	18
2.10 CONSIDERAÇÕES FINAIS.....	20
CAPITULO 3	21
3. METODOLOGIAS PARA AVALIAÇÃO DE IMÓVEIS.....	21
3.1 INTRODUÇÃO	21
3.2 REGRESSÃO LINEAR MÚLTIPLA.....	21
3.3 REGRESSÃO ESPACIAL	26
3.4 REDES NEURAIAS ARTIFICIAIS.....	35
3.5 VARIOGRAMAS	51

3.6	CONSIDERAÇÕES FINAIS.....	53
CAPITULO 4		55
4.	ESTUDO DE CASO – CIDADE DE BELO HORIZONTE	55
4.1	INTRODUÇÃO.....	55
4.2	A CIDADE DE BELO HORIZONTE.....	56
4.3	A CIDADE DE BELO HORIZONTE E A RMBH.....	57
4.4	PESQUISA DE DADOS.....	59
4.5	CONSTRUÇÃO DAS VARIÁVEIS	60
4.6	TRATAMENTOS DOS DADOS.....	66
4.7	ANÁLISE DOS RESULTADOS OBTIDOS NAS MODELAGENS.....	84
CAPITULO 5		90
5.	CONCLUSÕES.....	90

Lista de Tabelas

TABELA 4.1	Área total, população e densidade demográfica, Brasil, MG, RMBH, BH – 2000	57
TABELA 4.2	Domicílios em Belo Horizonte por espécie e unidade de planejamento – 2000	57
TABELA 4.3	Variáveis utilizadas nas modelagens com Regressão e RNAs	60
TABELA 4.4	Resultados estatísticos do modelo de preços hedônicos com a Regressão Linear	67
TABELA 4.5	Diagnóstico da Dependência Espacial.....	71
TABELA 4.6	Resultados estatísticos do modelo de preços hedônicos com a Regressão Espacial ..	72
TABELA 4.7	Resultados da estimação com as RNAs	75
TABELA 4.8	Resultados da estimação com as RNAs Espaciais	79
TABELA 4.9	Dados selecionados para a validação cruzada	83
TABELA 4.10	Resíduos quadráticos médios do processo de validação	84
TABELA 4.11	Resultados do processo de validação	88

Lista de Figuras

FIGURA 3.1	Visualização gráfica da regressão linear simples.	23
FIGURA 3.2	Representação de um Neurônio biológico.	41
FIGURA 3.3	Representação de um neurônio artificial.	43
FIGURA 3.4	Representação esquemática de uma RNA <i>feed-forward</i>	44
FIGURA 3.5	Amostragem em duas dimensões.	52
FIGURA 3.6	Exemplo de semi-variograma.	53
FIGURA 4.1	Regiões administrativas de Belo Horizonte.	58
FIGURA 4.2	Dados coletados no Mercado de Belo Horizonte / MG.	59
FIGURA 4.3	Dados de mercado (latitude x longitude).	70
FIGURA 4.4	RNA com a inclusão da variável de defasagem espacial – WY4.	78

Lista de Siglas e Abreviaturas

ABNT:	Associação Brasileira de Normas Técnicas.
AVALIAR:	Simpósio Brasileiro de Engenharia de Avaliações.
BH:	Cidade de Belo Horizonte
COBREAP:	Congresso Brasileiro de Engenharia de Avaliações Perícias.
EASYKRIG:	Toolbox em MATLAB para modelagem de dados e construção de variogramas.
EDO/DEA:	Análise por Envoltória de Dados sobre Dupla Ótica.
IBAPE:	Instituto Brasileiro de Avaliações e Perícias.
IBGE:	Instituto Brasileiro de Geografia e Estatística.
IPEAD:	Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Minas Gerais – UFMG.
Km:	Quilometro.
LM:	Multiplicador de Lagrange.
Ln:	Função Logaritmo Neperiano.
M ² :	Metro Quadrado.
MCP:	Modelo de Neurônio proposto por McCulloch e Pitts.
MCRL:	Modelo Clássico de Regressão Linear.
MG:	Estado de Minas Gerais.
MQO:	Mínimos Quadrados Ordinários.
MLP:	Multi Layer Perceptron.
NNSYSID:	Toolbox em MATLAB de Redes Neurais Artificiais.
OBD:	Optimal Brain Demage.
OBS:	Optimal Brain Surgeon.

PMBH:	Prefeitura Municipal de Belo Horizonte.
RMBH:	Região Metropolitana de Belo Horizonte.
RNA:	Rede Neural Artificial.
RNAs:	Redes Neurais Artificiais.
R\$/m ² :	Unidade Monetária (Real) por metro quadrado.
UP:	Unidade de Planejamento.
UTM:	PROJEÇÃO UNIVERSAL TRANSVERSAL DE MERCATOR.

CAPÍTULO 1

1. INTRODUÇÃO

1.1 PRELIMINARES

O mercado imobiliário possui grande importância para a economia do país, tanto pela geração de empregos na indústria da construção civil, como pela arrecadação de tributos, impostos e taxas, muitas destas calculadas sobre o valor venal¹ dos imóveis. Diversos órgãos governamentais e privados utilizam o valor de mercado como parâmetro importante para as tomadas de decisões. O estudo deste mercado influencia diretamente, dentre outras análises, as elaborações de plantas genéricas de valores para cobrança de impostos e desapropriações, a cobrança de impostos sobre ganhos de capital pela Receita Federal, as decisões do Poder Judiciário, a determinação da garantia de operações dos agentes financeiros, as análises de viabilidade de empreendimentos e operações de compra e venda de imóveis.

Os preços dos imóveis podem ser decompostos nos preços de seus atributos ou de suas características intrínsecas e extrínsecas, que na economia urbana, são chamados de “modelos hedônicos”. Assim, formalmente na Engenharia de Avaliações, os preços dos imóveis têm sido definidos como a expressão monetária dos dados de mercado em oferta ou efetivamente transacionados, sendo representado por,

$$P = f(x_1, x_2, \dots, x_n, \beta) \quad (1.1)$$

¹ Valor mais provável pelo qual o imóvel será transacionado, calculado com base nos preços praticados no mercado imobiliário.

em que f é o indicativo da forma funcional, P é o preço do bem, x_1, x_2, \dots, x_n são as características ou atributos relacionados a questões estruturais (físicas), de localização e aspectos econômicos e temporais e β são os parâmetros a serem estimados. Com base em uma amostra de preços coletados no mercado imobiliário em estudo, o valor de mercado dos imóveis pode ser estimado e as avaliações têm sido realizadas utilizando as técnicas tradicionais da econometria, como a Regressão Linear, com o cálculo de seus coeficientes através do método dos Mínimos Quadrados Ordinários (MQO).

A aplicação da Regressão Linear exige o atendimento aos seus pressupostos básicos para que a análise estatística seja validada. Neste sentido, duas questões de grande importância têm sido discutidas no meio acadêmico, que são a autocorrelação espacial e o desconhecimento da forma funcional para o modelo a ser adotado. Em González e Formoso (2000) é feita uma análise conceitual das dificuldades encontradas na determinação da Regressão Linear, encorajando os estudiosos a buscar fundamentos em outras metodologias, como as Redes Neurais Artificiais (RNAs).

Como consequência advinda da dificuldade na aplicação da Regressão Linear na Engenharia de Avaliações, na revisão da NBR 5676², que culminou na NBR 14.653-2³, foram introduzidas novas metodologias, com a citação explícita dos tratamentos de dados pelas técnicas de modelagem das RNAs, da Regressão Espacial⁴ e da Análise da Envoltória de dados sob Dupla Ótica⁵ (EDO / DEA).

² Norma Brasileira da ABNT – Associação Brasileira de Normas Técnicas, sobre Avaliação de Imóveis Urbanos.

³ A NBR 14653, da ABNT, é composta de 7 partes, sendo a de número 2 referente à Avaliação de Imóveis Urbanos, substituindo a NBR 5676.

⁴ A Regressão Espacial foi introduzida no Brasil através dos trabalhos apresentados pelo Eng. Rubens Alves Dantas, durante o Congresso Brasileiro de Avaliações e Perícias – COBREAP, Dantas et al (2001).

⁵ O método EDO / DEA foi inicialmente introduzido no Brasil no Avaliar – Simpósio Brasileiro de Avaliações.

A utilização das RNAs na Engenharia de Avaliações tem sido ainda restrita, principalmente pelos resultados não animadores obtidos até o momento, relatados por diversos autores. Rossini (1997), investigando preços de imóveis no sul da Austrália, comparou os resultados obtidos com a RNAs e a Regressão Linear, indicando vantagens desta última abordagem em relação às RNAs, muito em função das dificuldades de implementação computacional para o treinamento da rede. Worzala et al. (1995), utilizando uma amostra de 288 dados de mercado, relataram que os resultados obtidos não foram satisfatórios, novamente considerando o tempo necessário ao treinamento da rede. Entretanto, outros autores obtiveram resultados satisfatórios com o emprego das RNAs, a exemplo de Tay e Ho (1994), que investigaram a aplicabilidade das RNAs para as avaliações em massa⁶, para a cidade de Singapura, utilizando o algoritmo *backpropagation*. No Brasil, estudos recentes têm demonstrado boa aceitação das RNAs, a exemplo de Guedes (1995), que propõe um estudo comparativo com a Regressão Linear, indicando um melhor desempenho das RNAs e Pelli e Braga (2005) que demonstraram o poder computacional das RNAs, dispensando artifícios matemáticos como as variáveis de interação.

Nos trabalhos desenvolvidos tendo como modelagem as RNAs não se identificou a inclusão de informações sobre a distribuição espacial dos dados de mercado. Entretanto, a econometria espacial, que investiga os efeitos espaciais presentes nos dados distribuídos no espaço urbano, tem tido um grande avanço nos últimos anos, principalmente com a utilização de recursos computacionais mais

⁶ Avaliação em massa – avaliação sistemática de grupo de imóveis, a uma determinada data, de uma mesma tipologia, em diferentes macrolocalizações no espaço urbano, apoiada pela utilização de procedimentos padronizados e testes estatísticos (González, 2003).

sofisticados (softwares⁷ e hardwares). Neste campo, a Regressão Espacial tem sido citada em trabalhos técnicos mais recentes, permitindo obter resultados superiores aos alcançados pela Regressão Linear. Em trabalho pioneiro nesta área no Brasil, Dantas et al (2001) introduzem e incorporam a questão espacial nas avaliações do mercado imobiliário. Neste estudo é estimado um modelo espacial para uma região da cidade do Recife, com uma amostra de apartamentos, distribuídos em quatro bairros e são encontradas fortes indicações de autocorrelação espacial. Em expansões deste trabalho, Dantas et al (2002a), Dantas et al (2002b) e Magalhães e Dantas (2002) encontram resultados mais consistentes, que reforçam a presença de efeitos de dependência espacial em dados imobiliários para a cidade do Recife.

A NBR 14653-2 cita explicitamente como metodologia científica a utilização da Análise da Envoltória de Dados (EDO/DEA). Esta modelagem surgiu como instrumento para avaliar a eficiência de um conjunto de unidades de produção, diferindo do enfoque estatístico tradicional na medida em que não se propõe a relacionar uma determinada medida com a média ou a mediana dos resultados das unidades comparáveis, mas sim com a unidade mais eficiente (benchmarking). Ao invés de ser um método com enfoque nas medidas de tendência central, a intenção é se situar nas unidades extremas que compõem a fronteira ou envoltória dos dados. A utilização dessa ferramenta para a engenharia de avaliações foi proposta pela primeira vez por Lyra (2002). Pelli e Moraes (2006) apresentaram uma nova metodologia para a avaliação de imóveis com a construção de um sistema híbrido através das RNAs e do EDO/DEA.

⁷ No mercado nacional existe o software SisPlanV (2005), que trata a regressão espacial conforme inicialmente proposta por Anselin (1998).

1.2 OBJETIVOS DO TRABALHO

Este trabalho pretende atingir dois objetivos: o primeiro, de caráter metodológico, que é demonstrar a possibilidade da utilização das RNAs nas avaliações em massa, incorporando os efeitos espaciais nas especificações dos modelos, como uma alternativa à utilização da Regressão Espacial ou Econometria Espacial, metodologia inicialmente introduzida por Anselin (1998). O segundo trata da pesquisa aplicada ao estudo do mercado imobiliário da cidade de Belo Horizonte, fornecendo subsídios a futuros pesquisadores para a elaboração de uma análise microeconômica do mercado habitacional, com a utilização dos tratamentos científicos através das RNAs, da Regressão Espacial e da Regressão Linear.

Neste trabalho serão comparados os resultados obtidos através das modelagens de dados para uma amostra colhida aleatoriamente no mercado imobiliário da cidade de Belo Horizonte / MG. Partindo-se do modelo de preços hedônico, desenvolve-se uma análise da formação dos preços da habitação, e com base nos dados coletados e nas variáveis selecionadas para a modelagem, será possível analisar diversos modelos e a influência de cada atributo no valor dos imóveis.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação foi dividida em 5 (cinco) capítulos, incluindo este que é a introdução.

Neste capítulo destacou-se a importância do mercado imobiliário e da determinação do valor de mercado dos imóveis. Mostrou-se que a literatura nacional é bastante carente com relação à abordagem das RNAs, e que o processo dos MQO

para o cálculo da regressão linear é o que tem sido mais difundido, embora em diversas situações pode levar o avaliador a cometer erros de estimação. Foi feita uma breve explanação sobre as novas metodologias para tratamento dos dados de mercado citadas na NBR 14.653, da ABNT, e embora o EDO/DEA tenha apresentado bons resultados para aplicação nesta área de estudo, neste trabalho serão abordados os tratamentos com a utilização da Regressão Linear, da Regressão Espacial e das RNAs.

No capítulo 2 será apresentado um breve resumo dos conceitos aplicados à Engenharia de Avaliações, ao mercado imobiliário, no segmento da habitação, e às técnicas de modelagem, com revisão bibliográfica, descrevendo os trabalhos mais relevantes que servirão de base teórica para as formulações empíricas realizadas nesta investigação do mercado imobiliário de Belo Horizonte / MG.

No capítulo 3 será feita uma revisão literária sobre as técnicas de modelagem aplicadas neste trabalho, particularmente aquelas que influenciaram diretamente os resultados obtidos.

No capítulo 4 serão disponibilizados os resultados obtidos, incluindo as tabelas, gráficos e figuras utilizadas, com a aplicação para o estudo de caso da cidade de Belo Horizonte / MG.

Como encerramento, o capítulo 5 conterà as conclusões e discussões sobre as metodologias, com a comparação direta dos resultados obtidos e relacionando as perspectivas futuras.

De forma geral, os melhores resultados foram obtidos com a aplicação das RNAs, pela sua capacidade de processamento de dados que se relacionam de forma não linear.

CAPÍTULO 2

2. MERCADO IMOBILIÁRIO – CONCEITOS E REVISÃO LITERÁRIA

2.1 INTRODUÇÃO

O mercado imobiliário, no segmento da habitação, por sua importância na economia nacional, tem sido foco de estudo por diversas entidades de caráter público e privado. No estudo deste mercado, a base para os cálculos de demanda habitacional, com foco na aplicação de recursos financeiros, é o valor de mercado, estimado através das avaliações em massa. A pesquisa de alternativas para as estimativas de valores justifica-se pela importância econômica e social deste mercado, evitando-se erros ou imprecisões indesejáveis nas mensurações realizadas nas atividades avaliatórias.

De forma geral, os modelos de avaliações em massa podem ser empregados, além dos estudos de demanda habitacional, também para a definição de planos diretores, para estudos de viabilidade econômica de novos empreendimentos imobiliários, e para cálculos para fins de tributação, especificamente o IPTU e o ITBI.

Este capítulo será focado na revisão literária da Engenharia de Avaliações, começando por um breve histórico, para em seguida descrever conceitualmente o mercado imobiliário, incluindo a revisão dos métodos e normas usualmente tomadas como base nos processos avaliatórios. Ainda neste capítulo será definido o método comparativo de dados de mercado, que é o mais utilizado nas avaliações de imóveis.

2.2 HISTÓRICO DA AVALIAÇÃO

Segundo os Anais do I Congresso de Engenharia de Avaliações, a atividade como especialidade pode ser considerada relativamente nova no Brasil. No início do século XX, foram publicados os primeiros trabalhos sobre avaliação de terrenos. Durante as décadas de 20 e 30, diversos trabalhos procuraram difundir a nova técnica, sendo destaques aqueles assinados pelos engenheiros Anhaia Melo, Lysandro Pereira, Ernani Nogueira e Luiz Carlos Berrini.

No final da década de 30, a utilização da estatística como ferramenta indispensável para a prática da engenharia de avaliações permitiu aos profissionais apresentarem estudos concretos, não só no que se referia ao valor médio, mas também quanto à distribuição amostral, saneamento de dados mercadológicos e intervalos de confiabilidade.

Os Congressos Brasileiros de Avaliações e Perícias (COBREAP) tiveram início em 1974 tornando-se referência nacional entre os avaliadores, proporcionando a discussão técnica dos trabalhos e a interação dos profissionais.

No I COBREAP, realizado pelo Instituto Brasileiro de Avaliações e Perícias (IBAPE), no ano de 1974, na cidade de São Paulo, foi apresentado o trabalho de Engenharia de Avaliações: Avaliações de Terras Conflagradas pelas Fraldas Urbanas propondo a utilização de Regressão Linear Múltipla e Inferência Estatística por influência do engenheiro Domingos Saboya, tornando-se precursor na área.

É importante ressaltar que as metodologias apresentadas influenciaram todos os trabalhos nos últimos 30 anos. Alguns engenheiros passaram a se dedicar exclusivamente à atividade no âmbito público e privado.

No Brasil o crescimento urbano verificado nos grandes centros conduziu o poder público a um grande número de desapropriações, o que obrigou a implementação de estudos técnico-científicos no campo das avaliações.

Em 1978, o avanço do setor imobiliário permitiu aos engenheiros que participavam das desapropriações acompanharem a evolução do mercado, em especial no que se referia às transações e locações.

Os IBAPE regionais começaram a se estender em todo o território nacional, com representação em cada estado. A partir destes, a Engenharia de Avaliações passou a contar com um novo suporte aos engenheiros que se especializavam na área, surgindo diversos seminários, cursos e simpósios visando a formação e reciclagem de engenheiros e peritos, com a finalidade de avaliar um bem.

Em 1989 foi publicada a primeira Norma Brasileira de Avaliações de Imóveis Urbanos – NBR 5676 / 89, pela Associação Brasileira de Normas técnicas (ABNT), definindo conceitos e metodologias a serem aplicadas nos trabalhos avaliatórios. Neste ano, durante o Governo Federal de Fernando Collor de Melo, ocorreu o processo de venda dos imóveis funcionais da União, em Brasília/DF. Esse trabalho foi desenvolvido sob a responsabilidade do corpo técnico de engenharia da Caixa Econômica Federal, cujo principal mérito foi utilizar a Inferência Estatística e a Regressão Linear em avaliações em massa, com reconhecimento de toda a comunidade nacional voltada às práticas da Engenharia de Avaliações. O fato foi fundamental para a conclusão do trabalho, que posteriormente teve grande impulso com a adoção sistemática de computadores no processo avaliatório. Atualmente a Engenharia de Avaliações está totalmente integrada com os sistemas

computacionais, sem os quais a qualidade técnica tão desejada nestes trabalhos não poderia ser alcançada.

Dantas (1998) afirma que a evolução dos processos avaliatórios e o abismo entre o técnico e o "leigo" vêm aumentando significativamente, culminando com a norma brasileira para Avaliação de Imóveis Urbanos - NB5676/89 da ABNT, que incorporou conceitos matemáticos estatísticos acessíveis somente a quem possuía a formação adequada. As metodologias avaliatórias foram incorporadas na norma técnica com o sentido de determinar, com um grau de precisão aceitável, quais os parâmetros mais significativos na formação do valor e como estes se relacionam.

Foi lançado em 1998 o AVALIAR⁸ - um evento eminentemente científico. Nesta ocasião, o destaque foi a qualidade dos temas abordados e de seus palestrantes.

O 1º. Congresso Internacional de Engenharia de Avaliações ocorreu na cidade de Valência/Espanha em 2002, onde a presença brasileira foi significativa, com a apresentação de trabalhos técnicos inovadores e de alta qualidade técnica.

Essa participação resultou em convênios firmados entre Universidades da Espanha e Institutos Brasileiros de Avaliações e Perícias.

O engenheiro Domingos Saboya, precursor da inferência estatística aplicada às Avaliações Imobiliárias no Brasil, previu as dificuldades em inserir, no meio científico, metodologias que contrapõem aos modos até então "habituais e fáceis de fazer avaliação". Em um artigo elaborado em 1989 e intitulado "A ESCOLA QUE VEIO PARA FICAR", Saboya descreve com virtude a dificuldade que é a introdução de novas verdades, pois em seu texto nos mostra que as verdades de

⁸ Avaliar – Simpósio Brasileiro de Engenharia de Avaliações

antes nunca foram mentiras, mas sim que o avanço tecnológico e científico leva-nos a novas tendências, calcadas em experiências anteriores, mas com comprovação científica.

Os cursos atualmente ministrados nas Universidades, nos Institutos voltados para as Avaliações e mesmo aqueles ministrados pelas empresas privadas têm sido focados nos fundamentos da Inferência Estatística e da Regressão Linear. Embora as estimativas de valores elaboradas com base na regressão linear múltipla sejam consistentes, normalmente podem apresentar inconvenientes que resultam em uma diminuição na precisão de seu cálculo, indicando claramente a necessidade da busca de novas metodologias. Desse modo, os profissionais que atuam nesta área da engenharia têm buscado constantemente métodos e técnicas que garantam uma conclusão segura nos trabalhos avaliatórios.

2.3 LEIS E REGULAMENTAÇÕES

A Avaliação de Imóveis Urbanos é atribuição legal de engenheiros civis e arquitetos, em decorrência da Lei Federal nº 5914 (art. 7º, item c), de 24/12/66, que organiza o exercício da profissão para Engenheiros e Arquitetos. Está atribuída a estes profissionais a competência exclusiva de avaliação de imóveis, máquinas e equipamentos. A resolução nº 205 de 30 de setembro de 1971 introduz o Código de Ética Profissional do engenheiro. As atribuições de cada categoria profissional são definidas pela resolução nº 218, de 29/06/1973 e as atividades de “Engenharia de Avaliações e Perícias de Engenharia” estão regulamentadas na resolução nº 345 de 27 de julho de 1990.

2.4 MERCADO IMOBILIÁRIO

As metodologias utilizadas na avaliação de bens do mercado imobiliário serão descritas com detalhes no próximo capítulo, contudo se faz necessário o entendimento deste mercado e o seu funcionamento, bem como compreender os mecanismos existentes.

Como ponto de partida, é importante fixar os conceitos sobre o que é o mercado imobiliário e as diferenças com outros mercados de bens.

2.4.1 ENTENDENDO O MERCADO IMOBILIÁRIO

O mercado imobiliário pode ser dividido em segmentos, como o mercado de apartamentos, casas, lojas, escritórios, andar corrido, terrenos e glebas (urbanas ou rurais), armazéns, vagas de garagem, etc. Outra divisão é a do mercado para compra e venda ou para locação (Gonzalez, 2003).

O mercado de imóveis tem um comportamento distinto de outros mercados, tais como o mercado de automóveis, eletrodomésticos e outros desta natureza. Os principais fatores que distinguem os imóveis de outros bens são a vida útil elevada, a singularidade, a sua localização e fixação espacial e as interferências das leis municipais, estaduais e federais.

A vida útil elevada dos bens do mercado imobiliário pode propiciar dificuldades na mensuração de seu valor, pois este é influenciado pelas características físicas concernentes ao padrão de acabamentos e ao estado de conservação. Imóveis com a mesma idade real poderão estar em estados de conservação diferenciados, em virtude da manutenção predial executada ao longo do tempo e do padrão de seus acabamentos.

Diferentemente de outros bens nos quais as características intrínsecas e extrínsecas não são tão diferenciadas, os imóveis do mercado imobiliário são singulares. Por mais coincidentes que sejam as características de determinados imóveis, pelos menos a sua posição ou localização será diferente e não existe no mercado imobiliário um imóvel igual ao outro. Neste sentido, em muitas ocasiões, a determinação do valor de um imóvel não é tarefa trivial e requer a aplicação de conhecimentos científicos.

O mercado imobiliário se constitui em um mecanismo dinâmico, com transformações ao longo do tempo, sendo afetado por diversos fatores, sejam eles valorizantes ou desvalorizantes. A atuação simultânea e desordenada de diversos incorporadores, empreendedores, construtoras e do próprio poder público tem como consequência a mutação constante deste mercado, o que reflete diretamente nos valores pelos quais os imóveis são ofertados ou transacionados.

De forma geral percebe-se que o mercado imobiliário possui componentes importantes, interagindo constantemente, e que são responsáveis pela formação dos preços praticados para os imóveis.

2.4.2 COMPONENTES BÁSICOS

Os componentes básicos do mercado imobiliário são os bens levados a mercado, as partes interessadas na venda e as partes interessadas na compra.

O estudo estatístico do mercado imobiliário somente trará bons resultados quando houver um equilíbrio entre estes três componentes. Obviamente, a situação ideal é aquela onde exista abundância de informação, estando presentes no mercado muitos vendedores, muitos compradores e uma grande disponibilidade de bens de diversas fontes (Dantas 1998).

Ao engenheiro avaliador cabe a tarefa de descrever o mercado imobiliário, indicando o grau de equilíbrio entre seus componentes. Quanto mais perfeito for este equilíbrio, mais competitivo é o mercado, e mais justo será o preço pago pelo bem. Contudo, não existe um mercado de concorrência perfeita e equilibrada. Mas estudos estatísticos devem ser evitados quando se tratar de situações extremas, como o caso de monopólio ou oligopólio.

2.5 VALOR DE MERCADO, PREÇO E CUSTO

A NBR 14.653, na parte 1 – Procedimentos Gerais, no item 3.44, define Valor de Mercado como sendo:

“Quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente”.

Esta quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem não é necessariamente o preço pelo qual este bem será transacionado ou ofertado. O valor de mercado é o resultado de um processo matemático e/ou estatístico de modelagem de dados. Estes dados são obtidos através da coleta de informação sobre os preços pelos quais os imóveis com características semelhantes estão sendo negociados ou ofertados.

Portanto, preço e valor são referências distintas. Ao passo que “Valor de Mercado” se refere ao valor mais provável de um bem, o “Preço” reflete a quantia monetária pela qual um determinado bem está sendo ofertado ou transacionado. É extremamente comum o resultado de uma avaliação ser diferente do preço em oferta ou de transação. O que não é esperado é que esta diferença seja bastante

grande, pois neste caso existem provavelmente questões a serem abordadas que justificarão o valor adotado.

O custo de um bem também não refletirá o valor de mercado, pois da definição anterior de Valor de Mercado, nem sempre o valor mais provável pelo qual o bem será negociado coincidirá com o seu custo de produção. O valor de mercado pode ser inferior, igual ou superior ao custo de produção.

2.6 ENFOQUES BÁSICOS NA AVALIAÇÃO DE BENS

A identificação do valor de um bem pode ser realizada, em geral, com a utilização de três distintos enfoques básicos (González 2003):

- A renda, onde o valor do bem é identificado a partir da renda que ele pode gerar durante sua vida econômica.
- A comparação, que toma por base os preços de bens semelhantes praticados no mercado;
- O custo, cujas bases de cálculo são os gastos diretos e indiretos necessários à produção do bem;

Dentre os três enfoques, a comparação direta é o mais adequado e confiável para a identificação do valor de mercado, por sua simplicidade e por utilizar menos subjetividade, quando comparado ao método da renda, para se chegar a ele.

2.7 METODOLOGIAS APLICÁVEIS

A metodologia a ser aplicada para avaliação de um bem é decorrente da natureza do bem a ser avaliado, da finalidade da avaliação e da disponibilidade, qualidade e quantidade de informações colhidas no mercado.

Segundo a norma específica da ABNT⁹, são previstos diversos métodos para identificar o valor de um bem. Neste trabalho, o enfoque a ser empregado é o Método Comparativo Direto de Dados de Mercado.

2.8 O MÉTODO COMPARATIVO DE DADOS DE MERCADO

A parte 2 da NBR-14653 estabelece que deva ser priorizado o uso do método comparativo de dados de mercado. Na impossibilidade de se utilizar este método, pode-se optar por outro método que seja específico para a tipologia em estudo.

A decisão em consumir determinado bem, de forma geral, se dá através do método comparativo. Quanto maior o valor do bem a ser consumido, maior deve ser o rigor para avaliação, bem como quanto mais “apertada” a economia, mais necessário será efetivar uma transação com base em um laudo de avaliação.

Para medir o valor de mercado de um objeto utiliza-se, intuitivamente, a comparação do mesmo com outros objetos semelhantes e com valores conhecidos, procedimento denominado processo comparativo.

Ao comprar-se qualquer produto, tais como automóveis ou eletrodomésticos, primeiro verifica-se o preço solicitado por diversos fornecedores, depois formar-se um conceito sobre o preço médio praticado, para finalmente decidir-se sobre a aquisição, de acordo com os interesses, condições de pagamento e disponibilidade financeira.

Na utilização do processo comparativo busca-se inferir um valor que seja representativo para o objeto avaliado, tomando como base outros objetos que

⁹ NBR 14.653 – Avaliação de Bens

guardam semelhanças entre si, e que as diferenças que existam sejam pequenas ou desprezíveis. Como o conhecimento de todos os objetos (a população) disponíveis em determinado mercado é, normalmente, inacessível na sua totalidade, faz-se uso de amostras, cujos valores médios fornecem estimativas do valor médio para todos os objetos que compõe a população.

É evidente que, quanto mais homogênea a população investigada, mais homogênea será amostra. Assim, ao buscar-se o valor de um automóvel zero Km, de determinada marca, modelo e ano, ou de um televisor com marca e modelo definidos, é provável que as amostras colhidas contenham preços próximos entre si e próximos também da média aritmética de todas as amostras colhidas. Este fato ocorre em função dos custos de aquisição dos produtos junto aos fabricantes serem semelhantes. Outra questão importante é a facilidade na obtenção de amostra representativa junto ao mercado e que favorece sobremaneira a obtenção de conclusões confiáveis acerca da média populacional nestes casos.

Por outro lado, ao estimar o valor de mercado de um imóvel, pelo processo comparativo, o avaliador enfrenta dificuldades significativas, especialmente se for considerado que a população é muito heterogênea, gerando amostras também heterogêneas. Os produtos oferecidos não apresentam marcas ou modelos suficientemente padronizados para torná-los homogêneos. Além disso, não dependem diretamente dos custos de produção, estando muitas vezes ligados a fenômenos culturais, locais e socioeconômicos.

Em ambos os casos, as amostras apresentam variação em torno de sua média aritmética. A diferença é que nas amostras homogêneas de produtos industrializados esta variação é reduzida, enquanto que as amostras heterogêneas, que servem de base para medir o valor médio do mercado imobiliário, ao contrário,

demonstram geralmente uma variação elevada em torno de sua média aritmética. Este fato gera um alto grau de incerteza nas conclusões sobre a média populacional deste produto. Na realidade, qualquer amostra colhida ao acaso poderá conter dados com valores distantes da sua média aritmética. Estas diferenças entre os dados coletados e a média da amostra são função das diferenças físicas entre os dados, dos fatores socioeconômicos e da aleatoriedade do mercado.

A parcela referente à aleatoriedade, sempre presente em qualquer mercado, pode ser definida como uma subjetividade inerente ao próprio ser humano no momento de atribuir um preço ao produto que deseja vender, ou aceitar o preço de um produto na hora de comprar. Sendo assim, ela não pode ser medida e se compõe dos erros ou resíduos não explicados de uma amostra colhida ao acaso (Dantas 1998).

A parcela referente às diferenças físicas entre os dados é nula ou quase nula nas amostras de produtos industrializados, mas extremamente importante em amostras do mercado imobiliário, causando grande heterogeneidade nestas amostras. Estas diferenças físicas são função das características intrínsecas e extrínsecas dos imóveis. Para que se possa medir estas diferenças, são coletados dados no mercado imobiliário com características assemelhadas. Portanto, para utilização deste método é indispensável a existência de um conjunto de dados que possam ser comparados. A comparação será feita com base nas características intrínsecas e extrínsecas, que são descritas pelas variáveis de entrada, explicativas ou independentes (NBR 14653).

2.9 CONSTRUÇÃO DAS VARIÁVEIS

As variáveis são representações numéricas das características intrínsecas e extrínsecas dos imóveis. É importante observar a relação existente entre as

variáveis selecionadas, no intuito de verificar a dependências ou não entre as mesmas.

Na Engenharia de Avaliações considera-se como variável dependente ou de saída o preço praticado no mercado (oferta ou transação) e, como variáveis independentes ou de entrada, as respectivas características físicas (atributos tais como área, frente, padrão, vagas de garagem), de localização (índice fiscal, setor urbano e a distância a pólos de influência), e temporais (a data de ocorrência do evento).

A variável dependente poderá ser especificada com base no preço total ou no preço unitário, usualmente em unidades monetárias medidas por m² de área. Porém existem outras possibilidades de uso tais como volume (quando o pé direito é diferenciado e representativo), preço por dormitório, ou por metro linear de testada.

Esta escolha é definida durante a análise dos dados coletados e é função dos modelos escolhidos para representar o mercado imobiliário. A escolha das variáveis independentes está diretamente ligada à diversidade de características, tanto intrínsecas quanto extrínsecas, dos dados pesquisados e ao comportamento do mercado imobiliário de cada região. Portanto, torna-se imprescindível que ao definir a priori quais as variáveis independentes a serem utilizadas, deve-se observar quais delas efetivamente influenciam e explicam a variação dos preços coletados. As variáveis independentes podem ser divididas basicamente em quatro grupos: quantitativas, qualitativas, proxy¹⁰ e dicotômicas (também conhecidas com binárias ou dummy) (NBR 14653).

¹⁰ Conceito introduzido na Engenharia de Avaliações pela NBR 14653: “Variável utilizada para substituir outra de difícil mensuração e que se presume guardar com ela relação de pertinência” (NBR 14653-2).

2.10 CONSIDERAÇÕES FINAIS

O conhecimento do mercado de imobiliário, bem com a sua análise qualitativa e quantitativa, tem papel decisivo no desenvolvimento das avaliações imobiliárias. A análise geral do mercado permite ao avaliador guiar o seu trabalho de forma científica, relacionando os atributos mais relevantes ao estudo, minimizando os eventuais erros de estimativa.

Existem na literatura brasileira, especializada na Engenharia de Avaliações, poucos livros ou materiais didáticos que sejam focados no estudo do mercado imobiliário, detalhando a construção das variáveis e na definição de métodos de pesquisa e amostragem, de forma a permitir a utilização de técnicas mais avançadas, como os algoritmos de agrupamento de dados. Em Gonzalez (2003), foi utilizado o algoritmo *K-means*¹¹ para a segmentação dos dados, gerando diversos sub-modelos¹², obtendo resultados satisfatórios. Entretanto, normalmente os trabalhos são executados com base no conhecimento do mercado imobiliário e de sua estrutura, sem a padronização na pesquisa e levantamento de base de dados mais confiáveis.

O próximo passo será o estudo dos tratamentos disponíveis a serem aplicados aos dados de mercado para a determinação dos valores de mercado dos bens a serem avaliados.

¹¹ Algoritmo de divisão dos dados em grupos, proposto por MacQueen em 1967, requerendo, em uma parte do processo, o cálculo das médias.

¹² Os sub-modelos foram construídos utilizando o algoritmo *K-means*, resultando em modelos para mercado grandes, médios e pequenos, com base nos valores unitários dos imóveis.

CAPÍTULO 3

3. METODOLOGIAS PARA AVALIAÇÃO DE IMÓVEIS

3.1 INTRODUÇÃO

No capítulo 2 foram discutidos os conceitos relativos à Engenharia de Avaliações, bem como foi elaborada uma revisão literária sobre o mercado imobiliário. Neste capítulo será feita uma revisão dos principais conceitos relativos às metodologias mais utilizadas, iniciando pela Regressão Linear, em seguida abordando a Regressão Espacial para finalizar nos conceitos básicos das RNAs.

3.2 REGRESSÃO LINEAR MÚLTIPLA

As relações que podem ser descritas por um modelo de regressão linear múltipla são comuns no campo da Engenharia de Avaliações. A formulação mais simples para explicar o comportamento dos preços no mercado imobiliário, pela metodologia tradicional, é representada pela equação (3.1) (Kmenta, 1988).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (3.1)$$

Onde Y_i é a variável dependente ou explicada, X_{i1}, \dots, X_{ik} são as variáveis independentes ou explicativas, normalmente associadas às características físicas, de localização, e aos aspectos econômicos, β_0, \dots, β_k são denominados de parâmetros da população e ε_i são os erros aleatórios do modelo. A letra minúscula i refere-se à i -ésima observação e a segunda letra minúscula usada identifica a variável independente em questão.

A estimação dos parâmetros é feita por inferência estatística com base em uma amostra representativa do segmento de mercado em análise, realizada pelo método dos MQO. O vetor de coeficientes do modelo, também chamado de vetor de preços implícitos ou hedônicos, é obtido por (3.2) (Kmenta 1988).

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (3.2)$$

Constata-se que, da combinação linear dos preços implícitos de cada uma das características do imóvel e as respectivas quantidades demandadas, chega-se ao valor estimado de mercado do imóvel.

Para que os parâmetros inferidos no mercado, pelo método dos MQO, sejam não-tendenciosos, eficientes e consistentes¹³, alguns pressupostos sobre as variáveis independentes, os resíduos e a especificação do modelo devem ser atendidos: as variáveis independentes não devem conter nenhuma perturbação aleatória e não deve existir nenhuma relação linear exata ou quase exata entre as mesmas; os erros aleatórios satisfazem as hipóteses de variância constante (modelo homocedástico), normalidade e ausência de autocorrelação; e ainda que o modelo esteja corretamente especificado, ou seja, na sua composição estejam incluídas apenas variáveis explicativas relevantes, e a escala das variáveis qualitativas envolvidas seja adequadamente escolhida, com o objetivo de garantir a linearidade do modelo. Este modelo é denominado de Modelo Clássico de Regressão Linear – MCRL (Gujarati, 2000).

¹³ Não-tendenciosidade indica que a média de todas as possíveis médias de amostras extraídas da população coincide com a verdadeira média da população; por eficiência entende-se que o melhor estimador não tendencioso é o que apresenta a menor dispersão das médias estimadas em torno da verdadeira média (menor variância), e que a propriedade da consistência indica que na medida em que a amostra cresce, a sua média se aproxima do verdadeiro valor da média da população.

Em geral, quando se trabalha com dados de corte transversal não faz sentido testar a autocorrelação dos erros aleatórios, sendo este cuidado indispensável em dados de séries temporais.

3.2.1 REPRESENTAÇÃO GRÁFICA DA REGRESSÃO LINEAR

A representação gráfica de um modelo de regressão linear simples é uma reta que está à menor distância medida na vertical, dos pontos observados, dispostos em um plano formado por dois eixos cartesianos, sendo normalmente utilizado um eixo horizontal para variável independente e o eixo vertical para a variável dependente (FIGURA 3.1).

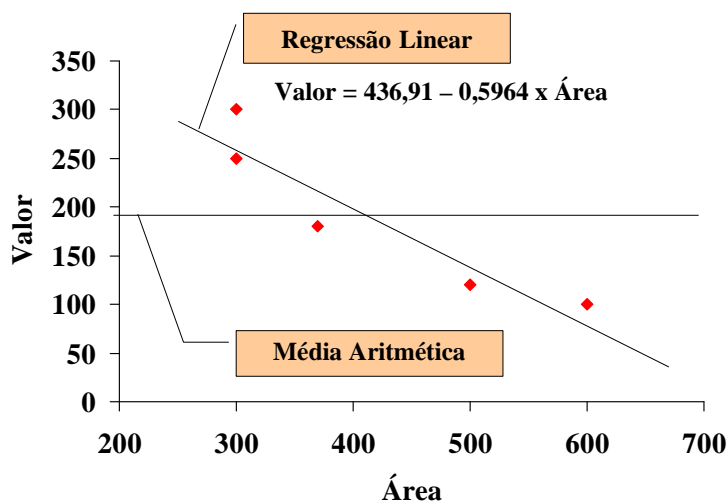


FIGURA 3.1 - Visualização gráfica da regressão linear simples.

Quando o modelo é composto por duas variáveis independentes, os pontos estão dispostos no espaço, formado por três eixos cartesianos, sendo um para a variável dependente, e um para cada variável independente. Cada eixo pode ser representado por um vetor. A situação ideal é aquela em que cada vetor seja independente, ou seja, o seu produto escalar é nulo.

Uma situação oposta seria aquela onde existe uma dependência linear perfeita entre as variáveis independentes, ocasionando a multicolinearidade entre seus eixos. Desta forma haveria a perda de uma dimensão do espaço, tornando-se impossível a estimação dos parâmetros da regressão.

Na prática, ocorre uma situação intermediária entre a multicolinearidade e a ausência total desta, devendo o avaliador investigar até que grau esta interferência entre as variáveis independentes torna-se prejudicial ao modelo.

3.2.2 HIPÓTESES BÁSICAS

Os modelos de Regressão Linear Múltipla estão sujeitos às hipóteses básicas e que se resumem, de forma geral, na verificação da normalidade, homocedasticidade, não auto-regressão e graus de colinearidade. Em Gonzalez (2003), se relacionam as causas das rupturas destes pressupostos básicos da regressão linear na Engenharia de Avaliações. Normalmente estas causas são devidas à distribuição espacial dos dados que provoca a autocorrelação espacial, ao desconhecimento da forma da linha de regressão, que supostamente pode assumir a forma não linear, à não normalidade dos resíduos e à multicolinearidade, com a influência simultânea de diversos atributos dos imóveis.

Em Kmenta (1978) ressalta-se o fato de que grande parte da econometria moderna deve sua existência à descoberta de que os estimadores dos MQO são, de fato, inconsistentes.

Em Costa Neto (2000) ressalta-se que, quando a linha da regressão não é conhecida de antemão, esta deve ser inferida juntamente com os parâmetros da regressão. No caso da regressão linear, as soluções obtidas com o método dos

MQO estão restritas aos modelos lineares em seus parâmetros, restringindo o leque das soluções viáveis aos problemas a serem estudados.

As principais dificuldades para a modelagem dos dados utilizando a regressão linear estão ligadas os seguintes fatores:

- A maioria dos modelos apresenta relações não lineares, cujo tratamento não atingiu ainda parâmetros satisfatórios de análise. Neste caso, modelos não lineares podem apresentar melhores resultados. Como artifício matemático é utilizado o procedimento de transformações nas variáveis, buscando linearizá-las;
- As amostras apresentam elevado nível de erros na mensuração das variáveis;
- O porte elevado da variação total dos valores em torno da média aritmética da amostra dificulta a definição de parâmetros para análise de resultados, resultando em modelos complexos.
- O alto grau de colinearidade entre variáveis independentes impossibilita a identificação da verdadeira influência de cada atributo no valor;
- A dificuldade na definição de escala numérica adequada para estudos de fatores subjetivos de influência, ligados a atitudes, expectativas, gosto dos consumidores.

Conclui-se que existem críticas à adoção exclusiva dos modelos de regressão linear na Engenharia de Avaliações e por estes motivos outras metodologias devem ser testadas.

3.3 REGRESSÃO ESPACIAL

3.3.1 INTRODUÇÃO

O valor de mercado de imóveis urbanos sendo estimado pelo MCRL, requer admitir-se a independência espacial das informações extraídas do mercado. Contudo, dados associados à posição que ocupam no espaço urbano (cidades, regiões homogêneas, bairros, pólos valorizantes e desvalorizantes), estão caracterizados pela dependência ou heterogeneidade espacial (Anselin, 1988). A questão principal é que as observações levantadas no mercado apresentam indexação no espaço, tendo como característica a continuidade, com uma variação gradual de valores na vizinhança. Na presença destes efeitos, os resultados obtidos pelo MCRL não são capazes de explicar com fidelidade o comportamento do mercado imobiliário, podendo gerar avaliações tendenciosas, inconsistentes ou ineficientes (Dantas 2002). Para corrigir estas anomalias recomenda-se a aplicação da metodologia denominada Econometria Espacial, que usa como ferramenta estatística a Inferência Espacial. Esta metodologia foi desenvolvida inicialmente por Matheron (1965) e recebeu grande impulso nas ampliações realizadas por Anselin (1988), principalmente na parte aplicada, com o desenvolvimento da ferramenta computacional denominada SpaceStat (Anselin, 1990).

Dantas (2001) reafirma que a literatura internacional reconhece a importância da questão espacial na avaliação de imóveis e tem tratado o problema de diversas formas, e em muitos casos com a utilização de variogramas, que são úteis para testar os efeitos de dependência espacial.

No Brasil, o primeiro trabalho com uso da metodologia definida por Anselin (1998) foi elaborado por Dantas *et al* (2001), representando a incorporação

da questão espacial à avaliação de imóveis. Tal estudo estimou um modelo espacial para uma região da cidade do Recife, com uma amostra de apartamentos situados em 59 edifícios residenciais, distribuídos em quatro bairros e encontra indicações de autocorrelação espacial. Em expansões deste trabalho, Dantas *et al* (2002a), Dantas *et al* (2002b) e Magalhães e Dantas (2002), com ampliação da amostra e do número de bairros, encontram resultados mais consistentes, que reforçam a presença de efeitos de dependência espacial em dados imobiliários na cidade do Recife.

Nos MCRL usualmente empregados nas avaliações de imóveis é comum a inclusão de variáveis indicativas da macrolocalização dos imóveis, tais como a distância a pólos de influência e a definição de regiões homogêneas de valores. Contudo, a microlocalização, que leva em conta a interação espacial entre os dados, normalmente não tem sido considerada. Quando os dados estão distribuídos espacialmente, como é o caso de imóveis no mercado habitacional, podem existir erros de medidas em relação à exata localização do imóvel, como também efeitos de interações espaciais (González 2003). Por estas razões deve ser considerado um fator adicional ao modelo tradicionalmente adotado, que é a autocorrelação ou dependência espacial. A não consideração deste efeito, como vem ocorrendo rotineiramente na análise do comportamento do mercado imobiliário, pode gerar problemas de estimação, pois, na presença de autocorrelação espacial nos resíduos, os parâmetros estimados por (3.2) são ineficientes (Dantas 2001). Neste caso, testes de hipóteses e os intervalos de confiança inferidos, não são mais válidos e as decisões tomadas com base neles são enganosas (Dantas 2001). Assim, a dependência espacial dos preços observados em relação aos preços dos imóveis vizinhos provocará estimações tendenciosas e inconsistentes para os parâmetros, em virtude de um erro de especificação no modelo, pela não inclusão de uma

variável dependente espacialmente defasada no modelo (3.1). Em ambos os casos, o MCRL mostra-se inadequado, devendo ser substituído pelos Modelos Espaciais, estimados por uma nova metodologia denominada Modelagem por Econometria Espacial.

3.3.2 MODELAGEM POR ECONOMETRIA ESPACIAL

Anselin (1998) afirma que existem dois tipos de efeitos que podem ser encontrados nos dados distribuídos espacialmente: o efeito de heterogeneidade espacial e o efeito de autocorrelação ou dependência espacial. O primeiro diz respeito à instabilidade dos parâmetros em relação à macro região em que se situam os dados e, na ausência de dependência espacial, podem ser tratados pela metodologia tradicional ou pelas RNAs; o segundo efeito diz respeito a uma interação espacial entre os dados, que pode afetar o termo de erro, a variável dependente ou ambos. Neste caso, a econometria espacial é capaz de realizar estimações seguras dos parâmetros do modelo. Os efeitos de autocorrelação espacial no termo erro devem ser tratados pelos Modelos de Erros Espaciais, através da inclusão de um fator de defasagem espacial nos erros aleatórios do modelo (3.1) e que será apresentado na seção 3.3.3, enquanto que os efeitos de dependência entre os preços de cada imóvel e os preços dos imóveis vizinhos devem ser tratados pelos Modelos de Defasagem Espacial, onde se inclui uma variável dependente espacialmente defasada, como variável explicativa no modelo (3.1), que será mostrado na seção 3.3.4.

Existem duas maneiras de se diagnosticar a presença de efeitos de dependência espacial em uma amostra: pela análise gráfica do variograma ou utilizando-se testes estatísticos específicos como os testes de Moran I e os testes

LM¹⁴ Robusto (erro) e LM Robusto (defasagem). No primeiro caso, a inferência espacial é realizada pelo processo denominado de Krigeagem¹⁵, desenvolvido por Matheron (1965); no segundo caso, a modelagem espacial é realizada conforme a metodologia desenvolvida por Anselin (1988), que é apresentada a seguir e utilizada no estudo de caso desenvolvido no capítulo 4.

3.3.3 MODELAGEM PELA METODOLOGIA DESENVOLVIDA POR ANSELIN

Para diagnosticar a presença de efeitos de dependência espacial, bem como introduzir estes efeitos no modelo (3.1), pela metodologia desenvolvida por Anselin, é necessário definir, previamente, uma matriz de pesos espaciais, conhecida como W . No caso mais simples, W é uma matriz simétrica em que cada elemento w_{ij} , é igual a 1 (um) se i e j são vizinhos e igual a zero no caso contrário. Por convenção, os elementos diagonais são iguais a zero, ou seja, $w_{ii} = 0$. Outras matrizes, como as propostas por (Cliff e Ord, 1981) e (Case et al, 1993), consideram a importância dos vizinhos através de uma ponderação correspondente ao inverso da distância ou ao inverso do quadrado da distância entre eles. Normalmente, esta distância é calculada com base nas coordenadas geográficas dos imóveis que compõe a amostra. Cuidado especial deve ser dado quando existirem, na amostra, imóveis localizados em um mesmo edifício. Nesta circunstância, a distância calculada utilizando as coordenadas geográficas será igual a zero, o que contradiz a lógica de mercado. Apartamentos situados em um mesmo edifício possuem uma alta correlação espacial e a distância entre estes imóveis deve ser medida na vertical.

Em geral, a matriz W é padronizada por linha, assumindo a nomenclatura W_s (Dantas 2001). Neste caso, cada elemento de W_s , representado por W_{sij} , é

¹⁴ LM é a sigla de Multiplicador de Lagrange.

¹⁵ Método de estimação (por interpolação) que considera a distancia entre os dados distribuídos espacialmente.

obtido dividindo-se W_{ij} pela soma dos elementos da linha i a que pertence, ou seja $w_{ij}^s = w_{ij} / \sum_j w_{ij}$. Nesta matriz, os elementos das linhas somam 1. Este procedimento, além de facilitar a interpretação dos pesos, como uma média ponderada dos valores dos vizinhos, assegura a compatibilidade entre os modelos (Anselin e Bera, 1998). O argumento principal a favor do uso de uma matriz de peso espacial é que esta associa uma variável em certo ponto do espaço (preço dos imóveis para o mercado de habitação) às observações da mesma variável em outros lugares do espaço. Neste trabalho será utilizada, por simplicidade, a notação W para a matriz de pesos espaciais ponderada por linha, calculada com base no inverso da distância entre os imóveis.

Os principais testes utilizados para detectar a autocorrelação espacial são Moran I, LM Robusto (erro) e LM Robusto (defasagem). O teste de Moran I é o mais usado nos estudos de dados de corte transversal de unidades geográficas. O problema deste teste é que ele não identifica o tipo de efeito (erro ou defasagem espacial). Por isso, serão utilizados testes mais específicos: o LM (erro) Robusto, para detectar efeitos de autocorrelação espacial no termo de erro; e o LM (defasagem) Robusto, para verificar a presença de efeitos de defasagem espacial na variável dependente. A seguir, estes testes serão apresentados de maneira resumida. Maiores detalhes podem ser encontrados em Anselin (1988a). É importante frisar que a validade destes testes exige a aceitação das hipóteses de normalidade e homocedasticidade dos resíduos de MQO, obtidos pelo modelo (3.1).

Teste LM Robusto (erro)

O teste LM (erro) Robusto é assintótico realizado a partir da estatística (3.3), que tem distribuição Qui-quadrado com um grau de liberdade, sob a hipótese

nula de não existência de autocorrelação espacial no termo erro. A estatística de teste é dada por:

$$\text{LM (erro)} = \frac{[e'We/(s^2/n)]^2}{[tr(W^2 + W'W)]} \stackrel{a}{\approx} \chi_{(1)}^2, \quad (3.3)$$

onde e é o vetor de resíduos de MQO, W a matriz de pesos espaciais, $s^2 = e'e/n$ a estimativa de máxima verossimilhança da variância do modelo (3.1), n o número de dados da amostra e tr o operador denominado traço da matriz.

Assim, se a estatística de teste for superior ao ponto crítico da distribuição Qui-quadrado, com um grau de liberdade, rejeita-se a hipótese de não autocorrelação espacial nos resíduos do modelo clássico de regressão.

Teste LM Robusto (Defasagem)

O teste LM (defasagem) Robusto é também assintótico, realizado a partir da estatística (3.4), que tem distribuição Qui-quadrado com um grau de liberdade, sob a hipótese nula de não existência de defasagem espacial na variável dependente. A estatística de teste é dada por

$$\text{LM (defasagem)} = \frac{\{[e'Wy/(s^2)]^2\}}{\{(WXb)'MWXb/s^2 + tr[W^2 + W'W]\}} \stackrel{a}{\approx} \chi_{(1)}^2 \quad (3.4)$$

onde e é o vetor de resíduos de MQO, W a matriz de pesos espaciais, y o vetor de observações na variável dependente, $s^2 = e'e/n$ a estimativa de máxima verossimilhança da variância do modelo (3.1), X a matriz das variáveis independentes, b o vetor de parâmetros estimados via MQO; n o número de dados da amostra $M=I-X(X'X)^{-1}X'$ e tr o operador denominado traço da matriz.

A hipótese de não autocorrelação espacial na variável dependente do modelo clássico de regressão será rejeitada se a estatística de teste for superior ao ponto crítico da distribuição Qui-quadrado com um grau de liberdade.

Uma vez detectada a presença de autocorrelação espacial nos dados, faz-se necessário introduzir extensões no modelo tradicional representado na equação (3.1), considerando-se os efeitos da autocorrelação espacial nos erros, pelo Modelo de Erro Espacial, e os efeitos ocasionados pelas interações entre os preços, pelo Modelo de Defasagem Espacial, como será mostrado a seguir.

3.3.4 O MODELO DE ERRO ESPACIAL

A autocorrelação espacial no termo de erro está relacionada a erros de medida ocasionados pelas divisões artificiais das unidades geográficas, como os limites estabelecidos para os bairros ou regiões consideradas homogêneas de uma cidade, que não necessariamente coincidem com a realidade estudada. Isto é, na prática, o consumidor não tem o conhecimento exato dos limites que dividem os bairros ou regiões. No mercado habitacional há uma tendência de efeito de transbordamento de um bairro de maior importância sobre os seus vizinhos. Por exemplo, o bairro de Floresta, em Belo Horizonte, devido à sua importância no contexto urbano e à grande demanda por habitação, foi se estendendo sobre os bairros Santa Tereza e Colégio Batista. Outro fator que pode gerar a autocorrelação espacial nos erros é a omissão de variáveis locais relevantes, notadamente as variáveis de microlocalização.

Para tratar adequadamente este tipo de efeito espacial nos dados, a primeira modificação com relação à equação (3.1) será considerar o processo espacial autoregressivo no termo de erro, da seguinte forma:

$$\varepsilon = \lambda \mathbf{W}\varepsilon + u \quad \text{ou} \quad \varepsilon = (\mathbf{I} - \lambda \mathbf{W})^{-1} u, \quad (3.5)$$

onde λ representa o coeficiente de autocorrelação espacial do termo erro; u é normalmente distribuído com média zero e variância constante; \mathbf{I} é a matriz identidade e \mathbf{W} a matriz de pesos espaciais ponderada. Substituindo (3.5) em (3.1) resulta no seguinte modelo de erro espacial:

$$Y = X\beta + (\mathbf{I} - \lambda \mathbf{W})^{-1} u \quad (3.6)$$

Para estimações eficientes dos parâmetros do modelo (3.6) é necessário usar o estimador de verossimilhança¹⁶, que consiste em maximizar a função de log-verossimilhança¹⁷ dada por (3.7), utilizando-se de técnicas de otimização não linear.

$$L = \frac{n}{2} \ln(\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \lambda \mathbf{W}| - \frac{1}{2\sigma^2} \varepsilon' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \varepsilon, \quad (3.7)$$

onde n representa o número de dados da amostra, \ln o símbolo do logaritmo natural, σ^2 a variância do modelo e as demais variáveis têm a mesma definição da equação (3.4).

Como comentado na seção 3.3.3, quando os erros são autocorrelacionados espacialmente, os parâmetros estimados pelo modelo (3.1) são não eficientes, isto é, os desvios padrões que se encontram associados a eles são tendenciosos. Assim, os testes de hipóteses e os intervalos de confiança construídos não são mais válidos e os resultados obtidos a partir deles são enganosos.

¹⁶ Os estimadores de máxima verossimilhança são considerados como sendo aqueles valores dos parâmetros que geram, com maior frequência, a amostra observada (Kmenta 1988).

¹⁷ Como os valores dos parâmetros que maximizam a função de verossimilhança (1) são os mesmos que maximizam seu logaritmo, então se pode operar com $L = \ln(l)$ (Kmenta 1988).

3.3.5 O MODELO DE DEFASAGEM ESPACIAL

O efeito de defasagem espacial é ocasionado pela dependência espacial criada como consequência da interação espacial entre os preços dos imóveis, conhecido como “efeito de vizinhança” (Dantas 2001). Quando um comprador e um vendedor realizam a transação de um imóvel, eles não somente levam em consideração as suas características estruturais e locacionais, mas também são influenciados pelos preços dos imóveis vizinhos. Neste caso, esta influência é medida pela inclusão de uma variável adicional no modelo (3.1), dada por $W \times Y$, sendo W a matriz de pesos espaciais e Y o vetor de preços dos imóveis, que é a variável dependente espacialmente defasada (Anselin 1998). Cada elemento WY_i , do vetor WY é formado por uma ponderação dos preços dos imóveis vizinhos. Esta variável serve também para captar os efeitos de dependência espacial não considerados explicitamente nas variáveis locacionais comumente utilizadas, como questões ligadas à segurança, saúde e educação (Dantas 2001). A introdução do termo de defasagem espacial, como variável explicativa, serve como “proxy”¹⁸ para as variáveis independentes omitidas que estão correlacionadas com as características locacionais (Pace, Barry e Sirmams, 1998). Com a incorporação desta variável, o modelo (3.1) passa a ser

$$Y = X\beta + \rho WY + \varepsilon, \quad (3.8)$$

onde ρ é o coeficiente de autocorrelação espacial da variável WY , ε é uma variável aleatória independente e identicamente distribuída.

¹⁸ Variável utilizada para substituir outra de difícil mensuração e que se presume guardar com ela relação de pertinência (NBR 14653-2).

Tendo em vista que a variável WY é aleatória, a estimação por MQO não é adequada, porque viola um dos pressupostos básicos do MCRL¹⁹. Observe-se também que, ao comparar os modelos (3.1) com (3.8), constata-se no primeiro a falta da variável WY , o que gera um grave erro de especificação²⁰. Neste caso, as avaliações realizadas por (3.1) são tendenciosas e inconsistentes. Da mesma forma que no modelo (3.6), a estimação deve ser realizada pelo método da máxima verossimilhança, que consiste na maximização da função (3.9) utilizando técnicas de otimização não linear.

$$L = \frac{n}{2} \ln(\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \rho \mathbf{W}| - \frac{1}{2\sigma^2} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} \quad (3.9)$$

3.3.6 ESCOLHA DE MODELOS

Uma maneira de escolher o modelo a adotar – o Modelo de Erro espacial ou o Modelo de Defasagem Espacial - pode ser feita pela comparação do valor absoluto das estatísticas (3.3) e (3.4). Assim, quanto maior for o valor encontrado na estatística de teste, maior será o efeito espacial correspondente a esta estatística, conforme argumento de Anselin e Rey (1991).

3.4 REDES NEURAIIS ARTIFICIAIS

3.4.1 INTRODUÇÃO

Nesta seção procurou-se descrever com maiores detalhes a parte conceitual, com o objetivo de permitir uma utilização do texto como uma primeira

¹⁹ As variáveis explicativas não devem conter nenhuma perturbação aleatória.

²⁰ No modelo devem estar presentes todas as variáveis explicativas relevantes.

leitura para os profissionais de Engenharia de Avaliações que estejam ingressando nesta área.

Os estudos de RNAs são relativamente novos. De 1943, quando da publicação do artigo de autoria de Warren McCulloch²¹ e Walter Pitts²², até hoje, muitas pesquisas vêm sendo realizadas em todos os âmbitos da ciência, como a medicina, biologia e engenharia. Não existem livros técnicos que tratem com especificidade as RNAs e sua aplicação na Engenharia de Avaliações. Porém, pode-se destacar trabalhos apresentados em congressos dedicados a esta metodologia e voltados para a Engenharia de Avaliações. Alguns colegas de profissão (Guedes 95), que acreditaram e pesquisaram o assunto em exaustão, forneceram subsídios para a aplicação desta metodologia e hoje há como resultado a citação das RNAs como metodologia científica reconhecida na NBR 14653, Avaliação de Bens, Parte 2 – Imóveis Urbanos.

3.4.2 BREVE HISTÓRICO DAS REDES NEURAIS ARTIFICIAIS

As primeiras pesquisas sobre RNAs tiveram início em 1943, com a publicação do artigo *“A Logical Calculus of the Ideas Immanent in Nervous Activity”*, de autoria de Warren McCulloch e Walter Pitts. Neste artigo, eles estabeleceram as bases da neurocomputação, desenvolvendo procedimentos matemáticos similares ao funcionamento dos neurônios biológicos. Esta contribuição teve um caráter estritamente conceitual, já que os autores não sugeriram aplicações práticas para o seu trabalho, e nem os sistemas propostos por eles tinham a capacidade de aprender.

²¹ Warren McCulloch – Psiquiatra e Neuroanatomista.

²² Walter Pitts – Matemático.

Em 1949, Donald Hebb²³ deu um passo importante na história das RNAs, pois foi o primeiro a propor uma regra de modificação de pesos, criando um modelo de aprendizado. Hebb propôs que a conectividade do cérebro é continuamente modificada conforme o organismo vai aprendendo tarefas funcionais diferentes e que agrupamentos neurais são criados por tais modificações.

Nos anos 50 apareceram implementações de RNAs através de circuitos analógicos e, naquela época, acreditou-se que o caminho para o entendimento da inteligência humana havia sido descoberto. Nathaniel Rochester²⁴ (1956) desenvolveu uma simulação em computador do neurônio de McCulloch & Pitts, com regra de treinamento Hebbiana.

Frank Rosenblatt²⁵ (1957) desenvolveu o Perceptron, que tinha como objetivo o reconhecimento de padrões ópticos (modelo da visão humana). Em 1958²⁶ Rosenblatt introduziu o primeiro modelo de rede neural artificial, estabelecendo a base para a Inteligência Artificial.

Bernard Widrow²⁷ desenvolveu um novo tipo de elemento de processamento de RNAs chamado de Adaline, equipado com uma poderosa lei de aprendizado e que, assim como o Perceptron, ainda possui aplicabilidade na atualidade. Fundou a primeira empresa de circuitos neurais digitais, a Memistor Corporation²⁸.

Marvin Minsky²⁹ escreveu o livro Perceptron, onde demonstrava as limitações da Inteligência Artificial. Em uma rigorosa análise matemática ficou

²³ Donald Hebb – Biólogo – Estudava o comportamento dos animais

²⁴ Nathaniel Rochester – IBM produziu um dos primeiros programas de IA

²⁵ Frank Rosenblatt – Pesquisador Norte Americano – (1928-1969)

²⁶ The perceptron – A probabilistic model for information storage and organization in the brain.

²⁷ Bernard Widrow – Cientista – Criador do Adaline

²⁸ Primeira empresa de circuitos neurais digitais, que produziu o “*memistors*”, elemento similar aos transistores, mas que realizava o ajuste de pesos de uma RNA.

²⁹ Marvin Minsky – Americano - Professor da Universidade de Carnegie-Mellon. Um dos pioneiros nos estudos da Robótica.

comprovado o baixo poder computacional dos modelos neurais utilizados na época, levando as pesquisas neste campo a ficarem relegadas a poucos pesquisadores. Entre a década de 70 e início da década de 80, o período ficou conhecido como a “era perdida no campo de redes neurais artificiais”.

Nos anos 80 o interesse pela área retornou, devido, em grande parte, ao surgimento de novos modelos de RNAs, como o proposto por John Hopfield³⁰ e Teuvo Kohonen³¹. Finalmente em 1986, David Rumelhart³² desenvolve o algoritmo de *backpropagation*, ou retropropagação do erro. Foi proposta a sua utilização para a aprendizagem de máquina, e ficou demonstrado como implementar o algoritmo em sistemas computacionais. Além disso, nesta mesma época, ocorreu o surgimento de computadores mais rápidos e poderosos, facilitando a implementação das RNAs. Os engenheiros da computação forneceram os artefatos que tornaram possíveis as aplicações da inteligência artificial.

3.4.3 REDES NEURAIS ARTIFICIAIS NA ENGENHARIA DE AVALIAÇÕES

Por ser recente, esta metodologia ainda é hoje desconhecida pela maioria dos profissionais atuantes na Engenharia de Avaliações. Contudo, alguns pesquisadores já afirmavam a importância deste novo conceito, desenvolvendo pesquisas nesta área, sendo, aliás, responsáveis pelos avanços que culminaram com a aceitação das RNAs como metodologia científica descrita na NBR 14.653 – Avaliação de Bens, Imóveis Urbanos – Parte 2, onde consta no Item 8 - Procedimentos Metodológicos – subitem 8.2.1.4.3 – Tratamento Científico, a seguinte denominação: “*Quaisquer que sejam os modelos utilizados para inferir o*

³⁰ John Hopfield – físico e biólogo professor da Universidade de Princeton.

³¹ Teuvo Kohonen – Professor acadêmico da Finlândia, especializado no estudo de Memórias Associativas.

³² David Rumelhart – Estudioso da psicologia cognitiva.

comportamento do mercado e a formação de valores, devem ter seus pressupostos devidamente explicitados e testados. Quando necessário, devem ser intentadas medidas corretivas, com repercussão dos graus de fundamentação e precisão. Outras ferramentas analíticas, para a indução do comportamento do mercado, consideradas de interesse pelo engenheiro de avaliações, tais como Redes Neurais Artificiais, Regressão Espacial e Análise Envoltória de Dados, podem ser aplicadas, desde que devidamente justificadas do ponto de vista teórico, com inclusão de validação, quando pertinente". Alguns trabalhos científicos de pesquisadores nesta área, bem como títulos de trabalhos apresentados em congressos e outras reuniões de caráter técnico, podem ser vistos no *site* da Pelli Sistemas Engenharia Ltda³³.

3.4.4 CONCEITOS BÁSICOS

As RNAs foram desenvolvidas a partir de uma tentativa de reproduzir em computador um modelo que simule a estrutura e funcionamento do cérebro humano. Uma RNA é um sistema que tem capacidade computacional adquirida por meio de aprendizado e generalização (Braga, Carvalho e Ludemir 2000). O aprendizado está relacionado com a capacidade das RNAs de adaptarem seus parâmetros como consequência com a interação com o ambiente externo. A generalização, por sua vez, está associada à capacidade destas redes de fornecerem respostas consistentes para dados não apresentados durante a etapa de treinamento.

As RNAs caracterizam-se por possuírem elementos de processamento de estrutura bem simples, inspirados no funcionamento do neurônio biológico, com conexões entre estes elementos de processamento. Cada conexão na rede tem um peso associado e este peso representa a intensidade de interação ou acoplamento

³³ www.pellisistemas.com.br

entre os elementos de processamento e se a sua natureza é excitatória ou inibitória (Haykin 2001).

As RNAs utilizam estruturas neurais artificiais, em que o processamento e o armazenamento das informações são realizados de modo paralelo e distribuído, por elementos processadores de complexidade relativamente simples. Estes elementos podem ser dispostos em camadas responsáveis pelas entradas das informações (camada de entrada – correspondendo às variáveis independentes utilizadas no mercado imobiliário), pelo processamento destas informações (camada intermediária) e pela produção de resultados (camada de saída – que corresponde às variáveis dependentes, normalmente valor unitário ou valor total), para posterior generalização.

Um modelo neural biológico consiste em uma rede de células, relativamente autônomas, dotadas, individualmente, de capacidade de processamento limitada. As células são ligadas por conexões, cada uma com um peso associado, que corresponde à influência da célula no processamento do sinal de saída. Pesos positivos correspondem a fatores de reforço do sinal de entrada e pesos negativos correspondem a fatores de inibição (Braga, Carvalho e Ludemir 2000).

Os modelos geralmente apresentam um conjunto de células de entrada, por onde são passadas as informações para a rede e um conjunto de células de saída, que apresentam os sinais de saída da rede, e um conjunto de células intermediárias. O conjunto composto pelos neurônios possui uma capacidade bastante poderosa no processamento de informações. Conceitualmente, pode-se considerar que as RNAs são modelos matemáticos que se assemelham à estrutura

do cérebro humano e possuem capacidade de aprendizagem para posterior generalização³⁴.

3.4.5 O NEURÔNIO NATURAL

O sistema nervoso humano é responsável pela tomada de decisões e pela adaptação do organismo ao meio ambiente, sendo esta função realizada através de um aprendizado contínuo. Este sistema é constituído de células, responsáveis pelo seu funcionamento, denominadas de neurônios (FIGURA 3.2). O cérebro humano apresenta aproximadamente 10 bilhões de neurônios e cerca de 60 trilhões de conexões entre eles (Haykin, 2001).

Estas células recebem, geram e transmitem os estímulos que chegam ou partem do cérebro.

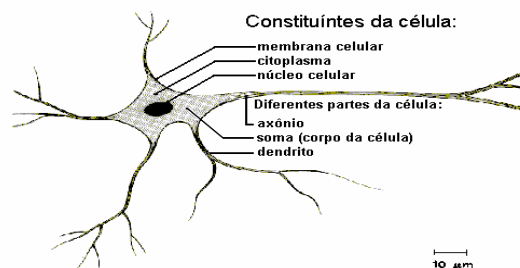


FIGURA 3.2 - Representação de um neurônio biológico

Fonte: apostila de RNA da Pelli Sistemas Ltda. (2005).

O neurônio é delimitado por uma fina membrana celular que possui determinadas propriedades, essenciais ao funcionamento da célula. A partir do corpo celular projetam-se extensões filamentosas, os dendritos, e o axônio (Braga, Carvalho e Ludemir 2000). Os neurônios são definidos como células polarizadas capazes de receber sinais em seus dendritos e transmitir informações por seus

³⁴ A generalização se refere ao fato de as RNAs produzirem saídas adequadas para entradas que não eram conhecidas durante o processo de aprendizagem.

axônios. Ao ser excitado, um neurônio transmite informações, através de impulsos, chamados potenciais de ação, para outros neurônios. Estes sinais são propagados como ondas pelo axônio da célula e convertidos para sinais químicos nas sinapses.

O neurônio biológico pode ser visto como o dispositivo computacional elementar do sistema nervoso, composto de muitas entradas e saídas. As entradas são formadas através das conexões sinápticas que conectam os dendritos aos axônios de outras células nervosas. Os sinais que chegam por estes axônios são pulsos elétricos conhecidos como impulsos nervosos ou potenciais de ação e constituem a informação que o neurônio processa para produzir como saída um impulso nervoso no seu axônio.

Dependendo dos sinais enviados pelos axônios, as sinapses podem ser excitatórias ou inibitórias. Uma conexão excitatória contribui para a formação de um impulso nervoso no axônio de saída, enquanto uma sinapse inibitória age no sentido contrário (Braga, Carvalho e Ludemir 2000).

A partir do conhecimento da estrutura e do comportamento dos neurônios naturais foram extraídas suas características fundamentais, utilizadas na criação de modelos de neurônios artificiais que simulam os reais. Estes neurônios artificiais são utilizados na formação das RNAs, se compondo em seus principais elementos de processamento.

3.4.6 O NEURÔNIO ARTIFICIAL – modelo MCP

O elemento básico que forma uma RNA é o neurônio artificial (FIGURA 3.3), conhecido também por nó ou elemento processador. Ele foi projetado por McCulloch e Pitts (Haykins 2001) e é baseado no funcionamento de um neurônio natural.

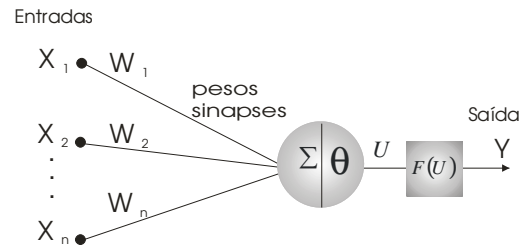


FIGURA 3.3 - Representação de um neurônio artificial

Fonte: Apostila de RNA da Pelli Sistemas Engenharia

O modelo do neurônio artificial proposto é bem simples. Ele possui n terminais de entrada x_1, x_2, \dots, x_n (que representam os dendritos) com pesos acoplados w_1, w_2, \dots, w_n a cada entrada, para emular o comportamento das sinapses. Alguns pesos possuem sinais excitatórios (+) e outros sinais inibitórios (-). Os valores de entrada e ativação dos neurônios podem ser discretos, nos conjuntos $\{0, 1\}$ ou $\{-1, 0, 1\}$ ou contínuos, normalmente compreendido nos intervalos $[0, 1]$ ou $[-1, 1]$.

Para cada uma das entradas x_i do neurônio da FIGURA 3.3 há um peso correspondente w_i . A saída linear u corresponde à soma das entradas x_i ponderadas pelos pesos correspondentes w_i , dada pela expressão (3.10):

$$u = \sum_i w_i x_i \quad (3.10)$$

A saída Y do neurônio é obtida pela aplicação de uma função $f(u)$ à saída linear u , indicada por (3.11):

$$Y = f(u) \quad (3.11)$$

onde f é chamada de função de ativação e pode assumir diversas formas lineares ou não lineares (Braga, Carvalho e Ludemir 2003).

3.4.7 REDES NEURAIS DE MÚLTIPLAS CAMADAS – MLP

A definição da arquitetura de uma RNA tem sua importância na medida em que restringe o tipo de problema que pode ser tratado. Uma RNA formada por um único elemento processador simples (neurônio artificial), como o apresentado na FIGURA 3.3, está limitada a solução de problemas linearmente separáveis. Existem diversos parâmetros que fazem parte da definição da arquitetura da rede, tais como o número de camadas da rede, número de neurônios em cada camada, tipo de conexão e a topologia da rede (Braga, Carvalho e Ludemir 2000).

Uma RNA é, portanto, formada por neurônios artificiais, onde cada neurônio possui capacidade limitada de processamento. Contudo, uma RNA, em função de sua arquitetura e topologia³⁵, pode apresentar boa capacidade computacional para a solução de problemas complexos. A FIGURA 3.4 representa uma RNA do tipo *feed-forward*, na qual cada neurônio executa uma função semelhante àquela da FIGURA 3.3.

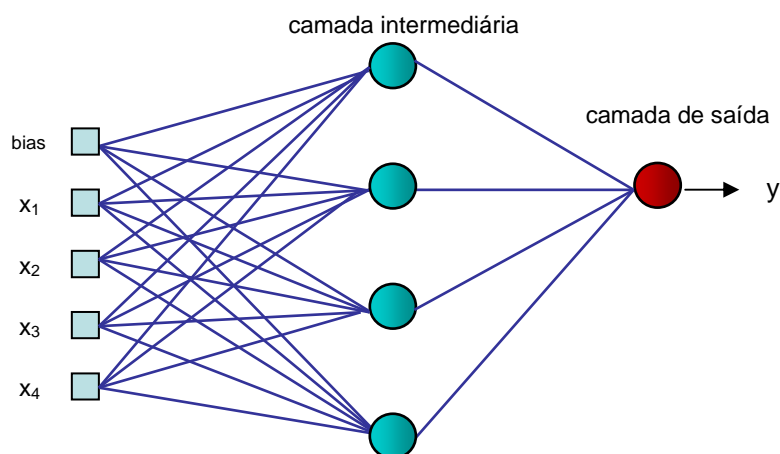


FIGURA 3.4 - Representação esquemática de uma RNA *feed-forward*.
Fonte: Apostila de RNA da Pelli Sistemas Engenharia.

³⁵ As topologias de RNAs mais conhecidas são a *feedforward*, onde a saída de um neurônio na *i*-ésima camada da rede não pode ser usada como entrada para os neurônios em camadas de índice menor ou igual a *i*, ao contrário da topologia *feedback*, que aceita como entrada de um neurônio a saída de outro neurônio localizado em uma camada de índice menor ou igual a *i*.

A estrutura apresentada possui quatro entradas x_1 , x_2 , x_3 , e x_4 e um *bias*³⁶, uma saída y e quatro neurônios na camada intermediária. Esta estrutura é capaz de resolver problemas de regressão, classificação ou predição (Braga, Carvalho e Ludemir 2003).

O número de entradas e saídas é em função da dimensão dos dados de entrada e saída, enquanto o número de neurônios nas camadas intermediárias depende da complexidade do problema, exigindo uma quantidade maior de neurônios para problemas mais complexos. Contudo, um número excessivo de neurônios na camada intermediária pode ter como consequência a obtenção de resultados indesejáveis, normalmente conhecidos como *overfitting*³⁷.

As funções utilizadas para o cálculo de ativação geralmente são não-lineares para garantir a plena funcionalidade das RNAs com múltiplas camadas de neurônios. As funções mais utilizadas são as que possuem um formato sigmoidal, tais como a sigmóide, a tangente hiperbólica, seno, gaussiana, etc.

3.4.8 APRENDIZADO DE REDES NEURAIS ARTIFICIAIS

As RNAs possuem a capacidade de aprender através da apresentação de exemplos. Os dados são apresentados nas entradas para que os parâmetros da RNA sejam ajustados de uma forma continuada, em função do processo de aprendizagem selecionado.

Para um determinado conjunto de dados (por exemplo, uma amostra selecionada aleatoriamente no mercado imobiliário), o algoritmo de aprendizado deve ser responsável pela alteração dos parâmetros da rede, para que em um

³⁶ Entrada com função semelhante ao intercepto (constante) da equação de regressão.

³⁷ Situação que ocorre quando a RNA está sobre parametrizada, possuindo mais neurônios do que os necessários ao treinamento.

número finito de iterações, haja convergência para uma solução (Braga, Carvalho e Ludemir 2003).

O critério de convergência será em função do algoritmo selecionado, existindo diversas implementações destes algoritmos. O objetivo do processo de aprendizado é a convergência para uma solução que será obtida através do ajuste do vetor de pesos w . De forma simplificada, o valor do vetor de pesos w na iteração $n + 1$ pode ser escrito como na função 3.12,

$$w(n + 1) = w(n) + \Delta w(n) \quad (3.12)$$

onde os algoritmos de aprendizado se diferem na maneira de obter o ajuste de Δw .

Normalmente, estes algoritmos são classificados como aprendizado supervisionado, não supervisionado e aprendizado por reforço.

No aprendizado supervisionado existe a presença de um professor ou supervisor, externo à rede, que tem a função de monitorar a resposta obtida para cada vetor de entrada. O conjunto de treinamento é formado por pares de dados de entrada e de saída (características dos imóveis da amostra e os preços coletados), onde se sabe, portanto, qual deve ser a resposta esperada da RNA. O ajuste de pesos é realizado de forma a obter na saída da rede o valor desejado para o imóvel, quando comparado com o preço praticado, dentro dos limites de tolerância inicialmente determinados. As RNAs do tipo MLP utilizam o aprendizado supervisionado.

O aprendizado não-supervisionado, como o próprio nome sugere, difere do aprendizado supervisionado, pela inexistência do professor ou supervisor. O aprendizado por reforço pode ser considerado como um meio termo entre os aprendizados supervisionado e não-supervisionado. Maiores detalhes podem ser

obtidos em Haykin 2001, Braga, Carvalho e Ludemir 2000, Braga, Carvalho e Ludemir 2003 e Kovács 2002.

3.4.9 DIFICULDADES NO APRENDIZADO E GENERALIZAÇÃO DE REDES MLP

O objetivo principal do processo de aprendizado é obter uma RNA com uma boa capacidade de generalização, tomando como base a amostra ou conjunto de dados coletados no mercado imobiliário.

No aprendizado supervisionado, com treinamento por correção de erros, o primeiro algoritmo de treinamento de redes MLP foi descrito em 1986 (Rumelhart, Hinton e Williams) sob a denominação de *backpropagation*. Este termo se deve ao fato de que o algoritmo se baseia na retropropagação dos erros para realizar os ajustes de pesos das camadas intermediárias (Haykin 2001). Desde então, diversos algoritmos foram propostos, destacando-se o algoritmo Marquardt³⁸ (Hagan e Menhaj 1994), que apresenta uma eficiência³⁹ bem superior no treinamento quando comparado com o treinamento com o algoritmo *backpropagation*.

Entretanto, a obtenção de um erro mínimo no processo de aprendizagem não garante a obtenção de valores de mercado consistentes para os imóveis a serem avaliados. Dentre os problemas conhecidos que dificultam a obtenção de uma boa generalização são o *overfitting* e o *underfitting*. O *overfitting* poderá ocorrer quando existir um excesso de neurônios na camada intermediária da rede, ou seja, na situação em que a RNA tem mais pesos do que necessário para a resolução do problema. O *underfitting*, por sua vez, ocorre quando a RNA possui menos

³⁸ O algoritmo foi proposto para ser incorporado ao treinamento *backpropagation*, baseado no processo de minimização do erro quadrático aplicado à regressão não linear.

³⁹ A eficiência está relacionada ao baixo número de iterações necessárias a convergência no treinamento, quando comparadas com o *backpropagation*, bem como na capacidade de tratar os mínimos locais normalmente existentes nas superfícies de erro (Haykin 2001).

parâmetros do que necessário. Neste caso, devem ser adicionados ao aprendizado, controles sobre o processo de treinamento e generalização, de forma a obter o ajuste ideal (Braga, Carvalho e Ludemir 2003). Existem diversas abordagens para solução destes problemas, entre elas estão os métodos construtivos e os métodos de poda. Os primeiros visam a construção gradual da RNA por meio da adição de neurônios na camada intermediária, até que o ponto ideal entre o treinamento e generalização seja alcançado. O processo se baseia na construção inicial de uma arquitetura com *underfitting*, e com a adição de neurônios, aproxima-se da arquitetura ideal. Os algoritmos de poda, por sua vez, percorrem o processo inverso, começando com uma estrutura inicial definida de forma empírica, mas visando a diminuição desta estrutura até a obtenção da arquitetura ideal. Os métodos de poda têm sido preferidos em relação aos métodos construtivos, muito em virtude dos algoritmos *Optimal Brain Damage* – OBD (Cun, Denker e Solla 1989) e *Optimal Brain Surgeon* – OBS (Hassibi e Stork 1993), descritos de forma resumida na próxima seção.

3.4.10 REDES NEURAIS COM “PODA”

A idéia básica deste método é iniciar a RNA com um número razoável de neurônios na camada intermediária e, durante a etapa de treinamento cortar as conexões (ou pesos) dos neurônios que possuem pouca influência no erro E. Neurônios que tiverem todas as conexões cortadas serão eliminados e, portanto, ao final dos “cortes”, sobrarão somente os neurônios realmente necessários à modelagem. A técnica de poda (Reed, 1993) reduz a complexidade da rede neural, melhorando sua capacidade de previsão, pois evita modelos sobre-parametrizados

(muitos neurônios e conexões) em que a possibilidade de sobreajuste (*overfitting*) é grande.

Existem basicamente dois métodos mais utilizados para a poda de RNAs: Optimal Brain Damage (OBD) e Optimal Brain Surgeon (OBS). Em ambos os métodos as conexões (ou pesos) são cortadas e a correspondente variação no erro E, chamada de saliência, é avaliada.

No método OBD as conexões são cortadas durante a etapa de treinamento e a RNAs não é retreinada após os cortes. No método OBS, as conexões são cortadas e, após o corte de uma conexão, a RNA é retreinada, permitindo que um número maior de cortes seja efetuado. Além disso, no método OBS a RNA é retreinada, aproximando-se os erros de treinamento por uma função quadrática, de modo a garantir a existência de um mínimo.

As técnicas de poda simplificam significativamente o processo de otimização da arquitetura e permite obter modelos com pequena possibilidade de sobreajuste (*overfitting*). Este fato pode ser observado comparando os resultados obtidos na determinação dos imóveis avaliados utilizando-se RNAs sem poda e com poda conforme será mostrado no próximo capítulo.

3.4.11 POLARIZAÇÃO E VARIÂNCIA

O ajuste da arquitetura das RNAs, com relação ao número de neurônios da camada intermediária, buscando evitar o *overfitting* ou o *underfitting*, conforme descrito de forma resumida na seção 3.4.9, tem sido caracterizado como o dilema entre a polarização e a variância (Braga, Carvalho e Ludemir 2003). RNAs com excesso de neurônios na camada intermediária tendem a ter uma maior variabilidade nas respostas (problema da variância), enquanto os modelos com um

baixo número de neurônios possuem baixa variância, mas geram respostas polarizadas, ou seja, são direcionadas para determinados resultados (Braga, Carvalho e Ludemir 2003).

Essas características das respostas obtidas nas saídas das RNAs são conflitantes, ou seja, a diminuição da polarização poderá levar a uma maior variância, bem como a diminuição da variância pode levar ao aumento da polarização. Para suavização do problema da variância nas saídas das RNAs, foi proposto a aplicação de algoritmos de *bagging*, que são métodos de geração de múltiplas versões de previsores e a utilização destes em uma árvore de decisão (Breiman 1994).

O processo inicia-se pela divisão do conjunto de dados, aleatoriamente, em um conjunto de teste T e um de aprendizagem L, sendo o tamanho de L bem superior ao de T. São construídos então conjuntos de dados L_B a partir do conjunto L utilizando-se de técnicas de amostragem. Os conjuntos L_B , são utilizados para o treinamento das RNAs. O conjunto de teste T é aplicado ao comitê de RNAs e a média dos resultados destas é comparada ao esperado em T gerando o erro médio quadrático e_B . O processo é repetido diversas vezes gerando um erro e_B médio.

O uso do *bagging* se torna atrativo quando se deseja projetar uma RNA cujo objetivo seja uma boa generalização (Haykin 2001), que é o caso da construção de RNAs para as avaliações em massa.

3.4.12 NORMALIZAÇÃO DOS DADOS DE ENTRADA

Como pré-processamento ao conjunto de dados para o treinamento é recomendável a normalização dos dados de entrada e de saída. Para as RNAs do tipo MLP um dos requisitos desejáveis é que os valores das entradas e das saídas

se encontrem no intervalo entre 0 e 1 para compatibilidade com a função de ativação de formato sigmoidal. Uma maneira de se proceder à normalização do conjunto de treinamento será descrita a seguir, que deve ser aplicada antes do início do treinamento das RNAs.

O primeiro passo é definir os limites mínimo ($l_{\text{mín}}$) e máximo ($l_{\text{máx}}$) do intervalo dentro do qual o conjunto de dados será normalizado e que tem como objetivo facilitar a convergência durante o algoritmo de treinamento da rede. Os dados são normalizados pela equação (3.13) e o retorno à escala original pela equação (3.14):

$$L_n = (L_o - L_{\text{mín}}) / (L_{\text{máx}} - L_{\text{mín}}) \quad (3.13)$$

$$L_o = L_n * L_{\text{máx}} + (1 - L_n) * L_{\text{mín}} \quad (3.14)$$

onde L_n é o valor normalizado, L_o o valor a normalizar, $L_{\text{mín}}$ e $L_{\text{máx}}$ são calculados respectivamente pelas equações (3.15) e (3.16).

$$L_{\text{mín}} = (4 \times \text{Limite}_{\text{inf}} - \text{Limite}_{\text{sup}}) / 3 \quad (3.15)$$

$$L_{\text{máx}} = (\text{Limite}_{\text{inf}} - l_{\text{máx}} \times L_{\text{mín}}) / l_{\text{mín}} \quad (3.16)$$

onde $\text{Limite}_{\text{inf}}$ e $\text{Limite}_{\text{sup}}$ são os valores mínimos e máximos respectivamente para cada variável do conjunto de dados para treinamento.

3.5 VARIOGRAMAS

O variograma é uma ferramenta básica de suporte às técnicas de krigeagem⁴⁰, que permite representar quantitativamente a variação de um fenômeno distribuído espacialmente (Huijbregts, 1975).

⁴⁰ Metodologia de inferência espacial desenvolvida inicialmente por Matheron (1965).

Considerando duas variáveis distribuídas espacialmente, P_1 e P_2 , onde $P_1 = Z(p)$ e $P_2 = Z(p+h)$, e que estas se referem ao mesmo atributo como, por exemplo, o preço do imóvel no espaço urbano, medido em duas posições diferentes, conforme ilustra a FIGURA 3.5 abaixo,

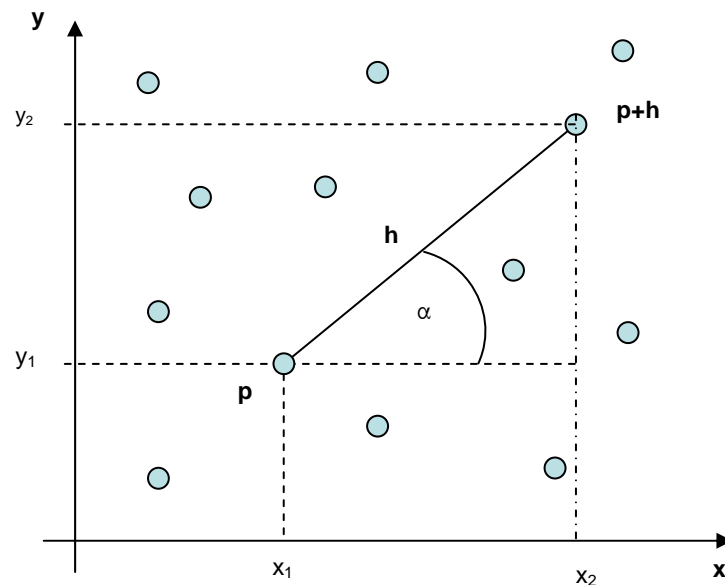


FIGURA 3.5 – Amostragem em duas dimensões.

onde p representa uma posição em duas dimensões, com componentes (x_i, y_i) , e h um vetor distância (módulo e direção) que separa os pontos, o nível de dependência espacial entre essas duas variáveis regionalizadas, P_1 e P_2 , é representado pelo variograma, $2\gamma(h)$, o qual é definido como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vetor distância h , dado por (3.17)

$$2\gamma(h) = E\{[Z(p) - Z(p+h)]^2\} \quad (3.17)$$

Para uma amostra de preços de imóveis $Z(p_i)$, $i=1, 2, \dots, n$, o variograma pode ser estimado por (3.18)

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(p_i) - Z(p_i + h)]^2 \quad (3.18)$$

onde $\hat{\gamma}(h)$ é o semi-variograma estimado, $N(h)$ é o número de pares de preços coletados $Z(p_i)$ e $Z(p_i+h)$, separados por um vetor distância \mathbf{h} , $z(p_i)$ e $z(p_i+h)$ são valores da i -ésima observação da variável regionalizada, coletados nos pontos \mathbf{x}_i e $\mathbf{x}_i+\mathbf{h}$ ($i = 1, \dots, n$), separados pelo vetor \mathbf{h} .

Um exemplo de semi-variograma é apresentado na FIGURA 3.6 abaixo,

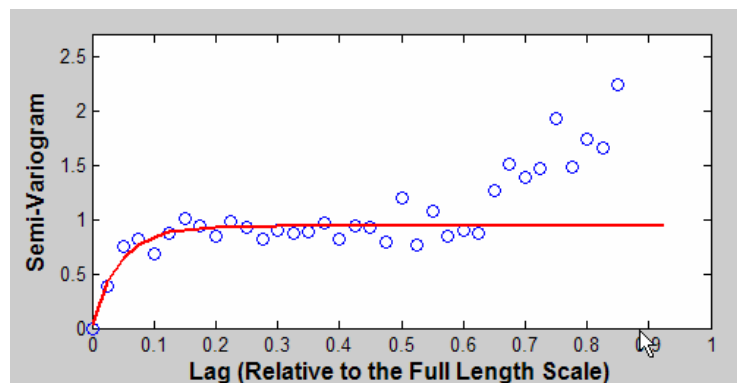


FIGURA 3.6 – Exemplo de semi-variograma.

onde em y está o valor de $\hat{\gamma}(h)$ para os pares de pontos separados pela distância h , representada no eixo x em valores percentuais de h .

3.6 CONSIDERAÇÕES FINAIS

Os processos de análise como são feitos atualmente, com a utilização da Regressão Linear Múltipla, não são satisfatórios, principalmente em função do desconhecimento da forma funcional do modelo, bem como pelo fato de não incorporar as questões relativas à autocorrelação espacial. A análise da regressão linear múltipla é uma técnica bastante usual, mas deve ser utilizada com bastante cuidado nas avaliações do mercado imobiliário.

Como alternativa, estudou-se a possibilidade da utilização da Econometria Espacial, que incorpora uma variável muito importante na formação dos preços dos imóveis, que é a variável de defasagem espacial. Portanto, espera-se que os modelos com a Regressão Espacial sejam superiores aos modelos com a Regressão Linear. As RNAs, pelo alto poder de processamento, e tendo em vista a eficiência dos algoritmos de otimização atuais, também é uma boa alternativa ao uso da Regressão Linear Múltipla, inclusive podendo à RNAs ser acrescida a variável de defasagem espacial, a ser abordada no próximo capítulo.

4. ESTUDO DE CASO – CIDADE DE BELO HORIZONTE

4.1 INTRODUÇÃO

No capítulo anterior foram apresentados os tratamentos científicos utilizados nesta dissertação, descritos na NBR 14.653-2, e que podem ser empregados para as modelagens de dados do mercado imobiliário. Neste capítulo destaca-se a cidade de Belo Horizonte como fonte do estudo empírico, com ênfase nas variáveis relacionadas com os preços de apartamentos na região metropolitana. Serão feitos alguns comentários sobre as características da cidade, com base nas informações do Censo Demográfico do IBGE (2000), visando descrever o espaço urbano em estudo.

Em seguida, como uma primeira abordagem, será construído um modelo econométrico com os dados fornecidos pela Caixa Econômica Federal, utilizando a Regressão Linear Múltipla, para em seguida proceder ao estudo da dependência espacial. Sendo o resultado positivo, o modelo espacial de preços hedônicos será estimado pela Regressão Espacial e as informações introduzidas nos modelos de RNAs.

Para o diagnóstico de dependência espacial serão utilizados semi-variogramas e testes de autocorrelação espacial, conforme metodologia definida por Anselin (1998). Para o tratamento dos dados serão utilizados os programas SisPlanV⁴¹, para a Regressão Espacial e a análise de autocorrelação espacial, o SisReN⁴² para a modelagem das RNAs, o NNSYSID20⁴³ (toolbox para Matlab) para

⁴¹ Sistema de Regressão Linear e Regressão Espacial desenvolvido pela Pelli Sistemas Engenharia Ltda.

⁴² Sistema de Regressão Linear e de Redes Neurais Artificiais desenvolvido pela Pelli Sistemas Engenharia Ltda.

⁴³ Toolbox para o Matlab, versão 6.5, que incorpora técnicas de poda para as Redes Neurais Artificiais.

o processo de poda das estruturas das RNAs e EASYKRIG⁴⁴ versão 3.0 (toolbox para Matlab) para a construção dos semi-variogramas.

4.2 A CIDADE DE BELO HORIZONTE

A cidade de Belo Horizonte, inicialmente chamada de "Cidade de Minas", foi inaugurada no dia 12 de dezembro de 1897 por Bias Fortes, presidente de Minas Gerais (1894-98). A primeira cidade planejada do país foi construída a partir de uma concepção urbanística elaborada pelo engenheiro paraense Aarão Reis. Ele queria enfatizar a modernidade e a desenhou prevendo separar os setores urbano e suburbano, delimitados pela Avenida do Contorno. Grandes avenidas, ruas largas e um parque central. Tudo que lembrasse Paris, Washington, e colocasse Belo Horizonte entre as grandes cidades do mundo. A realidade foi maior que o sonho e muitas previsões estavam equivocadas. A cidade cresceu além do esperado.

Inspirados por um belo horizonte que alimentava sonhos, os habitantes pediram ao Governo Provisório do Estado que mudasse oficialmente o nome "Cidade de Minas" para "Belo Horizonte". A mudança só ocorreu em 1906, através de um decreto expedido pelo então governador João Pinheiro da Silva.

A escolha de Belo Horizonte como capital do estado se deu principalmente por suas qualidades climáticas e topográficas, tendo sido comprovado à época que o terreno da cidade era seco e por este motivo não necessitava de prévia drenagem, facilitando a implantação das edificações. As condições de topografia e de solo se prestavam a um sistema perfeito de esgotos e águas pluviais. Em 17 de dezembro de 1893, Afonso Pena, na ocasião presidente de Minas Gerais (1892-94), promulgou a lei que designava Belo Horizonte para ser a

⁴⁴ Toolbok para Matlab, desenvolvido por Dezhang Chu e Woods Hole Oceanographic Institution.

capital do Estado. O prazo mínimo para a transferência definitiva do governo era de 4 anos, entretanto o tempo foi insuficiente e a cidade teve que ser inaugurada às pressas, ainda poeirenta e com prédios a construir. Sua consolidação levou anos.

4.3 A CIDADE DE BELO HORIZONTE E A RMBH

Os grandes déficits habitacionais do país estão concentrados nas regiões metropolitanas das grandes cidades. O IBGE – Instituto Brasileiro de Geografia e Estatística, no Censo Demográfico de 2000, contabilizou uma população de 2.238.526 para a cidade de BH, que corresponde a quase 50% da RMBH – Região Metropolitana de Belo Horizonte (32 municípios vizinhos a Belo Horizonte), conforme TABELA 4.1:

TABELA 4.1 - Área total, população e densidade demográfica
Brasil, MG, RMBH, BH – 2000

Especificação	População	Área (km ²)	Densidade
BH	2.238.526	330,90	6.764,96
RMBH	4.357.942	9.459,10	460,71
MG	17.891.494	586.552,40	30,50
BRASIL	169.799.170	8.514.215,30	19,94

Fonte: IBGE. Censo Demográfico 2000

Na TABELA 4.2 está indicada a quantidade de domicílios:

TABELA 4.2 - Domicílios em Belo Horizonte por espécie e unidade de planejamento - 2000

Nome UP	Total de domicílios	Domicílios particulares			Domicílios coletivos
		Total de domicílios particulares	Domicílios particulares permanentes	Domicílios particulares improvisados	
Noroeste	97.051	96.107	95.916	191	944
Norte	51.551	51.017	50.780	237	534
Oeste	77.902	77.112	76.949	163	790
Pampulha	40.440	39.768	39.668	100	672
Venda Nova	65.358	65.015	64.896	119	343
Barreiro	70.286	69.953	69.747	206	333
Centro-Sul	84.354	83.001	82.833	168	1.353
Leste	73.279	72.452	72.193	259	827
Nordeste	76.294	75.952	75.465	487	342
Belo Horizonte	636.515	630.377	628.447	1.930	6.138

Fonte: IBGE. Censo Demográfico 2000

A administração pública municipal está dividida nos setores listados na FIGURA 4.1:



FIGURA 4.1 – Regiões administrativas de Belo Horizonte

Fonte: PMBH (Prefeitura Municipal de Belo Horizonte)

Pelo porte da cidade de Belo Horizonte e considerando a quantidade de domicílios existentes, quantificados pelo censo de 2000 do IBGE, os modelos para avaliação em massa necessitam de grande robustez, com uma coleta de dados eficiente, buscando identificar os conglomerados urbanos, de forma a que as estimativas de valores sejam as mais próximas possíveis dos preços praticados.

Este é um desafio que tem enfrentado grandes instituições públicas, como a Caixa Econômica Federal e a Prefeitura Municipal de Belo Horizonte, para a construção e manutenção de modelos com fundamentação científica e que descrevam com eficiência o mercado imobiliário local.

4.4 PESQUISA DE DADOS

A pesquisa de dados no município de Belo Horizonte teve início em fevereiro de 2004 e término em setembro de 2005, perfazendo um período de 20 meses. Todos os dados coletados foram vistoriados, para identificação de seus atributos, com o preenchimento de ficha padrão de coleta de dados, que é fornecida junto com o software SisDat⁴⁵ – Banco de Dados Imobiliário. Estes dados foram coletados e fornecidos pela Caixa Econômica Federal. As coordenadas geográficas foram disponibilizadas pela Prefeitura Municipal de Belo Horizonte.

A partir da montagem dos mapas georreferenciados, foi possível identificar as coordenadas geográficas de cada elemento da base de dados, e com a utilização do software SisPlanV foi construída a matriz de pesos espaciais. Em seguida, o software calculou as distâncias no plano cartesiano, medidas em metros lineares, entre cada dado do banco de dados. Esta matriz de pesos foi utilizada na Regressão Espacial e nas RNAs. Os duzentos e cinquenta e cinco dados georreferenciados foram plotados através do software MapInfo, produzindo o mapa da FIGURA 4.2:

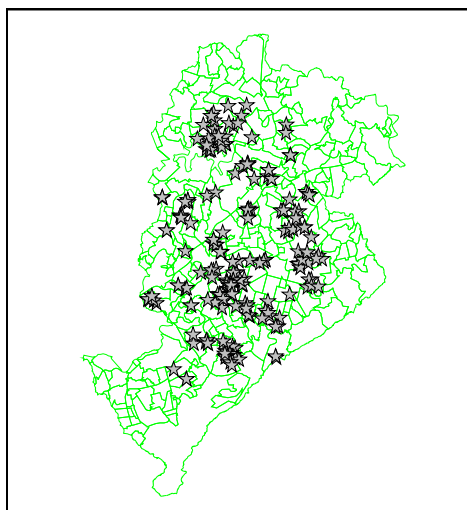


FIGURA 4.2 – Dados coletados no mercado imobiliário de Belo Horizonte / MG

⁴⁵ Software de banco de dados desenvolvido pela Pelli Sistemas Engenharia Ltda., para armazenagem de dados do mercado imobiliário.

4.5 CONSTRUÇÃO DAS VARIÁVEIS

Para a construção das variáveis independentes (de entrada), a serem utilizadas na modelagem de dados, optou-se por manter a estrutura existente no banco de dados fornecido pela Caixa Econômica Federal. Os valores numéricos das variáveis qualitativas deste banco de dados foram calculados com base em códigos alocados através da construção de escalas por meio do uso de dicotomias, em modelagens com o uso da Regressão Linear Múltipla. Como as variáveis serão as mesmas em todos os tratamentos científicos a serem aplicados aos dados, e considerando que um dos objetivos do trabalho é a comparação de resultados gerados por cada um destes tratamentos, optou-se por utilizá-las, sem alteração de suas escalas.

Após a obtenção dos dados brutos, foi utilizada a estatística descritiva para a exploração do conteúdo das variáveis numéricas. As medidas estatísticas da base de dados estão listadas na TABELA 4.3, que contém uma sumarização da amostra com a especificação das variáveis utilizadas na modelagem, apresentando os valores mínimos, médios, máximos e a amplitude.

TABELA 4.3 – Variáveis utilizadas nas modelagens com Regressão e RNAs

Variável	Mínimo	Média	Máximo	Amplitude
Setor Urbano	250	484	900	650
Nível / Elevador	5	8	26	21
Data Evento	1	5	20	19
Total de Vagas	0	121	400	400
Área Coberta	40	83	300	260
Área Descoberta	0	5,6	100	100
Dormitórios	2	2	4	2
Equipamentos	1	8	32	31
Unidades p/Acesso	2	11	80	78
Padrão Acabamentos	120	258	600	480
Estado de Conservação	240	473	600	360
Preços Unitários	400,00	960,96	3.333,33	2.933,33

A variável “Setor Urbano”, indicativa da macrolocalização dos imóveis, foi construída com base no índice fiscal da PMBH, ajustada a um domínio entre os

valores de 100 a 1000, ajuste este devido à defasagem dos índices fiscais com os preços dos imóveis no espaço urbano. A escala para a variável “Total de Vagas” foi construída com a utilização de variáveis dicotômicas e assume os valores zero, que indica a inexistência de vagas, 50 para as vagas descobertas, 70 para as vagas cobertas presas (necessitam de manobras) e 100 para as vagas cobertas e livres. Como exemplo, um apartamento que possua quatro vagas cobertas livres terá uma pontuação de 400 para a variável Vagas de Garagem. A variável Data do Evento é quantitativa, mensurada mês a mês, e indica a data de ocorrência do evento, assumindo o valor um para o primeiro mês da coleta (fevereiro de 2004). As variáveis “Padrão dos Acabamentos” e “Estado de Conservação” também tiveram as escalas ajustadas por meio do uso de dicotomias, procedimento que está em consonância com a NBR 14.653.

Os resultados apresentados na TABELA 4.3 indicam a existência de uma grande heterogeneidade nos dados, tanto do ponto de vista de localização, quando estrutural (física dos imóveis) e temporal (ocorrência do evento). Os preços de venda dos apartamentos variam de 400,00 R\$/m² a 3.333,33 R\$/m², que em preço total representa uma variação entre R\$ 25.000,00 a R\$ 1.000.000,00, ou seja, o mais caro tem preço 40 vezes superior ao mais barato. As áreas privativas interna dos imóveis também possuem grande variabilidade, sendo a maior área 7,5 vezes superior ao de menor área. Com relação à localização (variável setor urbano), também a variabilidade é bastante grande. Os imóveis mais bem localizados estão 3,6 vezes melhor pontuados em relação aos imóveis localizados nos setores urbanos mais populares. Observa-se também que a área descoberta dos apartamentos varia de 0,00 m² a 100 m², indicando uma variação de 100% entre os imóveis que não possuem área descoberta e aqueles que estão situados nos

pavimentos térreos, com áreas privativas descobertas. Existem também imóveis recém construídos e imóveis com idade de 40 anos, com conservação variando de ruim a boa e padrão de mínimo a luxo.

Em complementação à estatística descritiva inicial, foi realizada uma análise gráfica, buscando identificar a coerência na correlação das variáveis independentes em relação à variável dependente. Nos gráficos numerados de 4.1 a 4.21 observa-se a dispersão dos dados de cada variável independente em relação à variável dependente, bem como a distribuição de freqüências, que pode indicar se os elementos extremos estão deslocados dos demais ou se existe continuidade nos valores.

A reta existente nos gráficos das variáveis independentes versus valores unitários tem como finalidade fornecer um indicativo visual da correlação isolada (sem considerar o efeito do conjunto de variáveis nos modelos de regressão) entre cada variável independente e a variável dependente. Estas correlações, para os modelos de preços hedônicos do mercado de habitação, normalmente podem ser positivas ou negativas. A reta nos gráficos é calculada através da regressão linear simples entre a variável independente e a dependente. Outro objetivo da análise gráfica das variáveis é identificar a existência de dados discrepantes ou influenciantes para a variável independente. Os gráficos de números ímpares, exceto o gráfico 4.23, são de dispersão e os de número par são gráficos de distribuição de freqüências. Nos gráficos de dispersão, identifica-se que os dados com valores unitários acima de 2.000 R\$/m² são discrepantes para a maioria das variáveis. Como os dados, de forma geral, são extremamente heterogêneos, e buscando evitar a construção de modelos heterocedásticos, foram eliminados da modelagem todos os imóveis com preço unitário acima de R\$ 2.000,00 R\$/m².

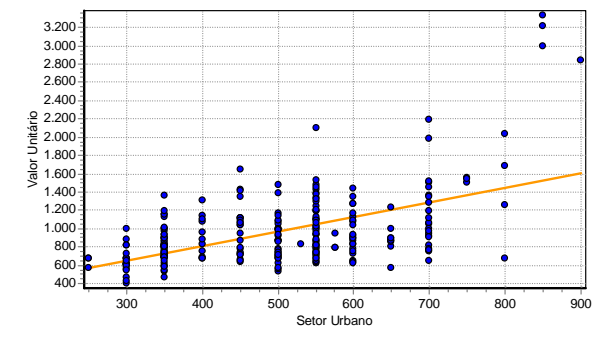


GRÁFICO 4.1 – Setor Urbano x Valor Unitário

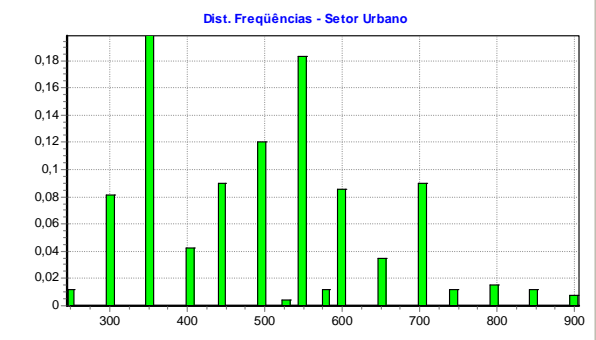


GRÁFICO 4.2 – Distribuição de frequência – Setor Urbano

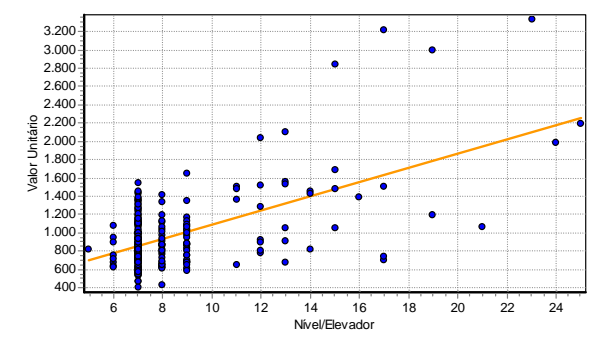


GRÁFICO 4.3 – Nível / Elevador x Valor Unitário

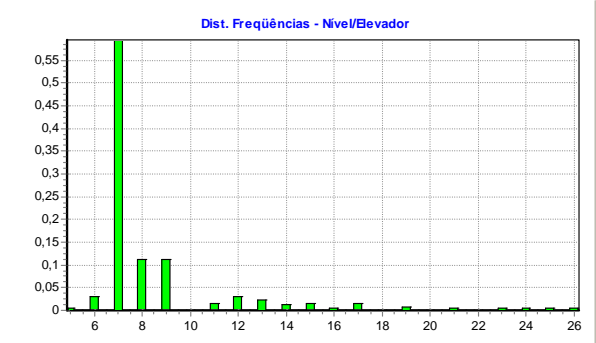


GRÁFICO 4.4 – Distribuição de frequência – Nível / Elevador

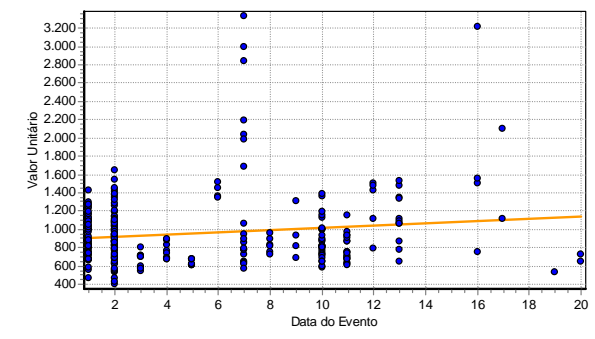


GRÁFICO 4.5 – Data Evento x Valor Unitário

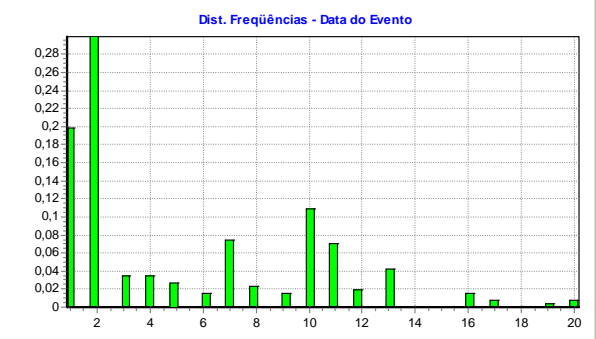


GRÁFICO 4.6 – Distribuição de frequência – Data Evento

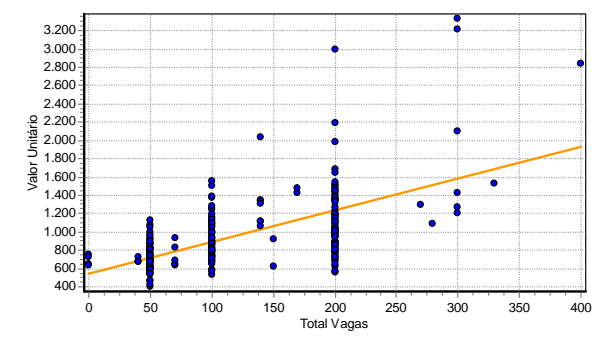


GRÁFICO 4.7 – Total de Vagas x valor unitário

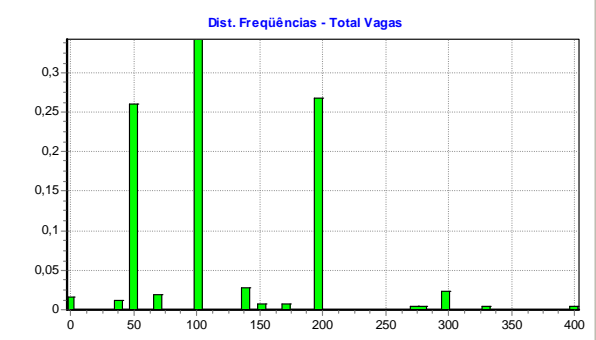


GRÁFICO 4.8 – Distribuição de frequência – Total de Vagas

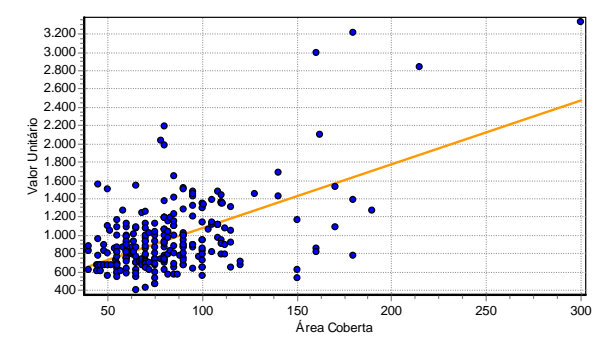


GRÁFICO 4.9 – Área Coberta x Valor Unitário

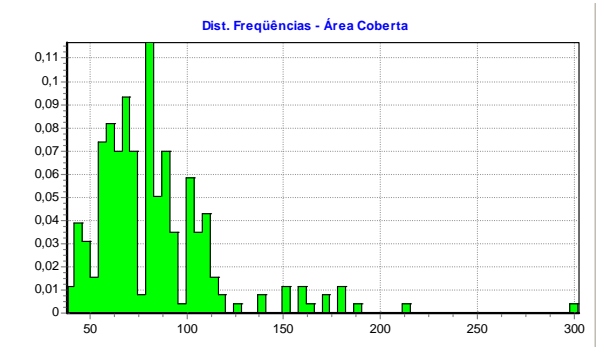


GRÁFICO 4.10 – Distribuição de frequência – Área Coberta

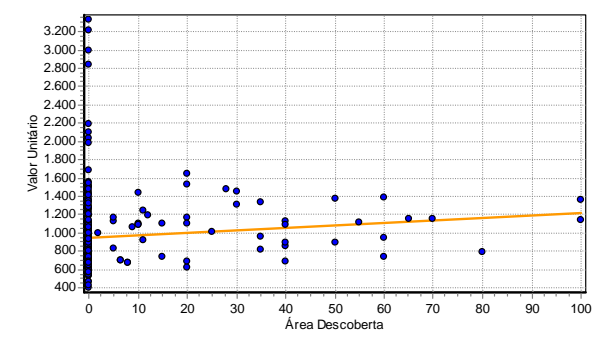


GRÁFICO 4.11 – Área Descoberta x Valor Unitário

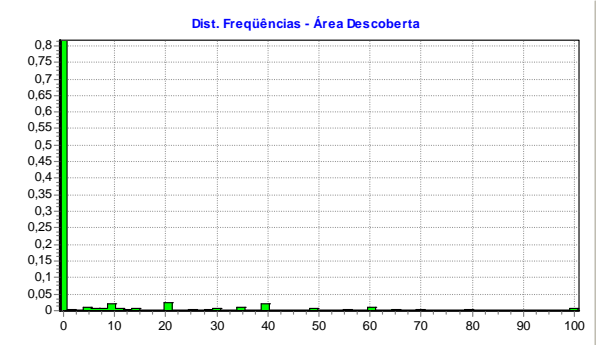


GRÁFICO 4.12 – Distribuição de frequência – Área Desc.

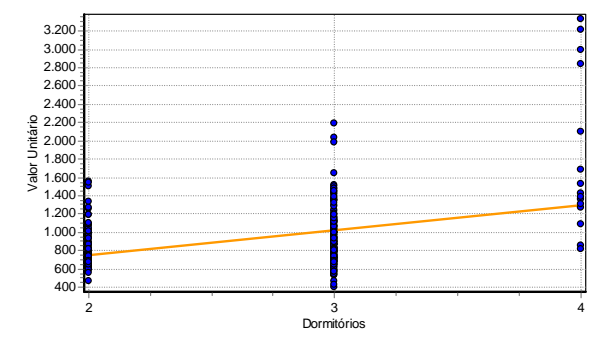


GRÁFICO 4.13 – Dormitórios x Valor Unitário

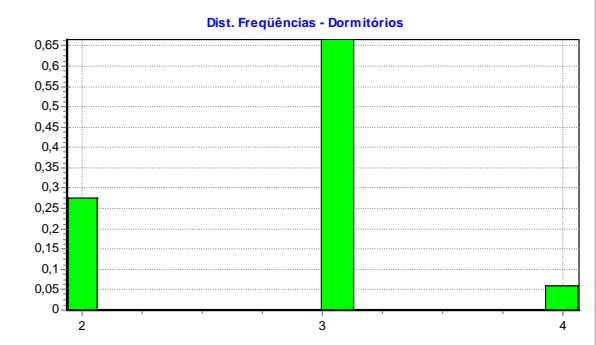


GRÁFICO 4.14 – Distribuição de frequência – Dormitórios

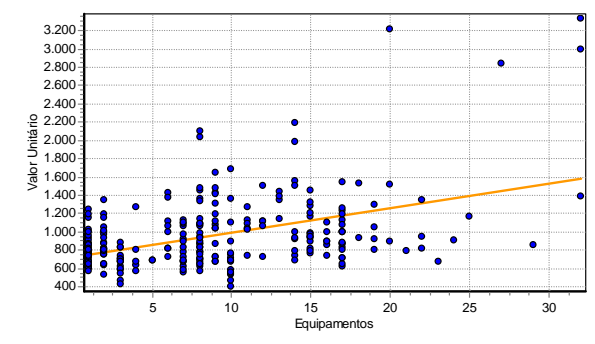


GRÁFICO 4.15 – Equipamentos x Valor Unitário

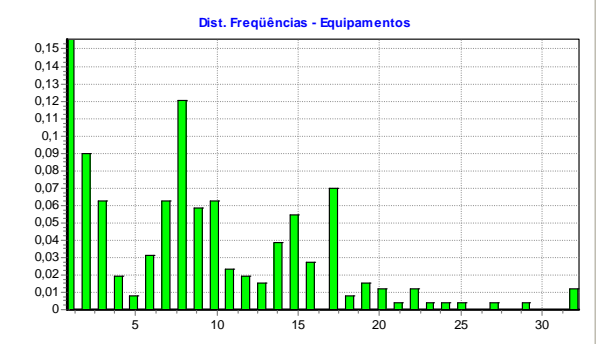


GRÁFICO 4.16 – Distribuição de frequência – Equipamentos

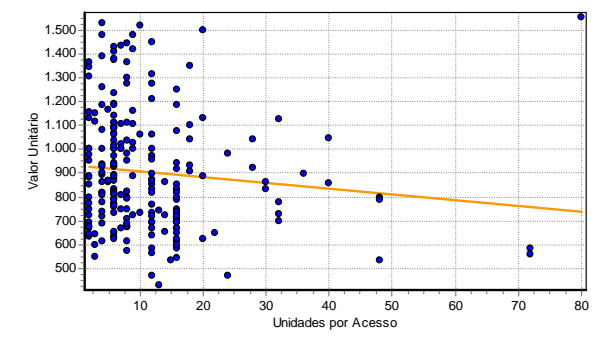


GRÁFICO 4.17 – Unidades p/Acesso x Valor Unitário

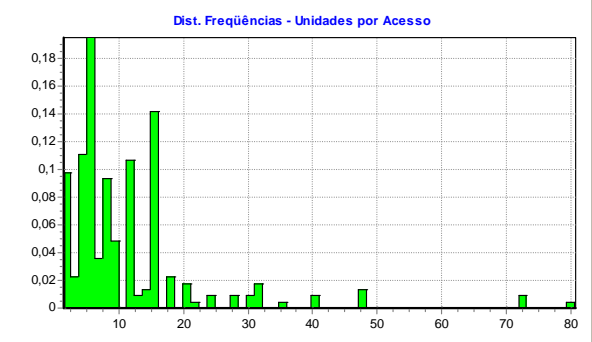


GRÁFICO 4.18 – Distribuição de frequência – Unid. Acesso.

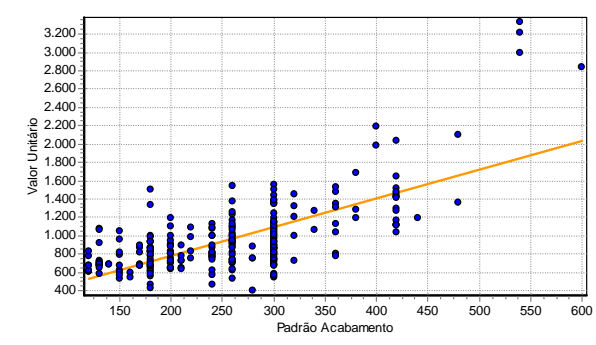


GRÁFICO 4.19 – Padrão Acab. x Valor Unitário

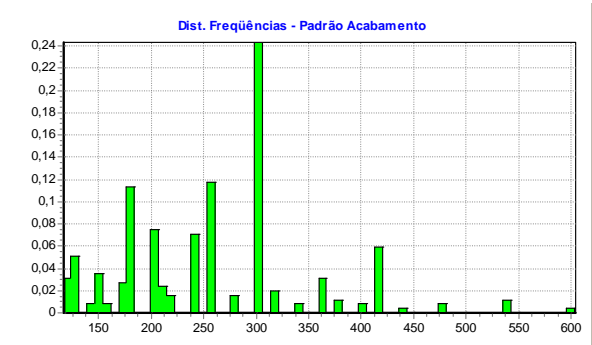


GRÁFICO 4.20 – Distribuição de frequência – Padrão Acab.

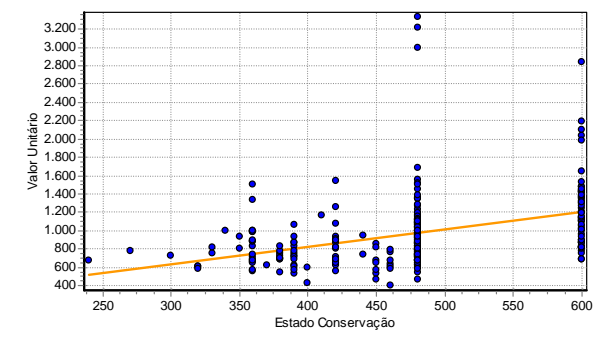


GRÁFICO 4.21 – Conservação x Valor Unitário

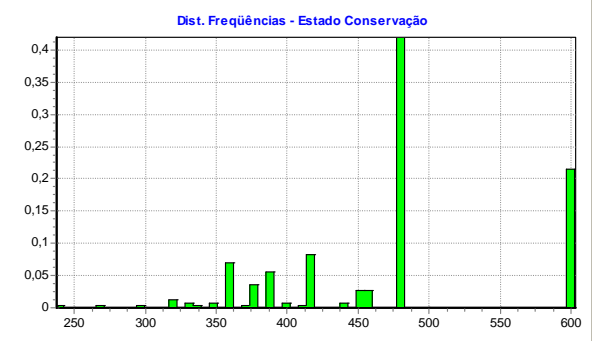


GRÁFICO 4.22 – Distribuição de frequência – Conservação

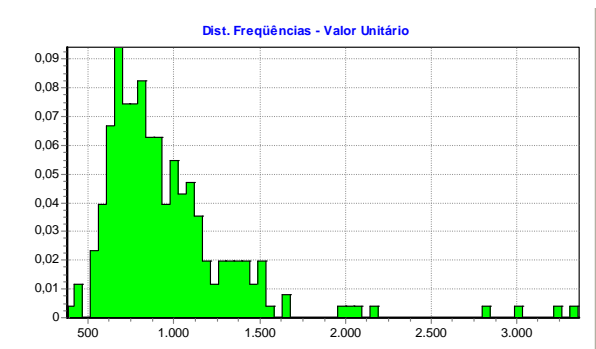


GRÁFICO 4.23 – Distribuição de frequência – Valor Unitário

4.6 TRATAMENTOS DOS DADOS

4.6.1 REGRESSÃO LINEAR

No modelo tradicional de preços hedônicos, o preço da habitação (P) é em função de seus atributos, notadamente aqueles referentes à localização (L), estruturais (E) e temporais (T), como definido na equação (4.1),

$$P = f(L, E, T, \beta) + \varepsilon \quad (4.1)$$

onde f é o indicativo da forma funcional, β são os parâmetros e ε são os resíduos ou erros aleatórios.

Para a estimação empírica do modelo (4.1) será utilizado, como uma primeira abordagem, o MCRL via MQO. Nesta etapa não será considerada a existência da autocorrelação espacial dos dados, cuja investigação será elaborada na próxima seção. Então, foi realizada uma regressão do preço de oferta ou transação sobre as variáveis independentes relacionadas na TABELA 4.3. Pelos resultados obtidos na TABELA 4.4 e no GRAFICO 4.24, verifica-se que os testes de hipóteses (estatística t) estão dentro dos parâmetros definidos pela NBR 14653 e que existe forte indicativo de normalidade dos erros aleatórios. Estes resultados foram inicialmente considerados consistentes, indicando nesta etapa que o modelo adotado é válido para a avaliação de apartamentos no mercado de Belo Horizonte. Durante o processo de modelagem, utilizando o método de “todas as regressões possíveis”, sem eliminação de variáveis, foram aplicadas as transformações \ln (logaritmo neperiano) e a função inversa, em todas as variáveis, buscando um modelo que apresentasse o maior poder de predição. Dentre os modelos disponibilizados foi selecionado o que apresentou o melhor coeficiente de correlação (C_C).

TABELA 4.4 – Resultados estatísticos do modelo de preços hedônicos com a Regressão Linear

Variáveis	Transformação	B (regressor)	t - Student	α (%) (bi caudal)
Setor Urbano	X	+0,0008226328371	8,24	< 0,01
Nível/Elevador	Ln(X)	+0,1613379329	3,50	0,05
Data do Evento	X	+0,005733353163	2,35	1,95
Total Vagas	X	+0,001348528869	6,70	< 0,01
Área Coberta	Ln(X)	-0,5026981916	-8,71	< 0,01
Área Descoberta	X	+0,002915663288	4,42	< 0,01
Dormitórios	X	+0,1168737474	4,70	< 0,01
Equipamentos	X	+0,009908174092	5,31	< 0,01
Unidades p/ Acesso	Ln(X)	-0,05386150912	-3,97	< 0,01
Padrão Acabamento	X	+0,001020533186	6,45	< 0,01
Estado Conservação	X	+0,0009130749008	5,99	< 0,01
Preços Unitários	Ln(Y)	+7,039344981		

O teste de hipótese, usando a estatística t, para todas as variáveis independentes, apresentou significância abaixo de 10%, e o f calculado (análise de variância) foi de 53,71 com a significância do modelo abaixo de 0,01%, limites recomendados pela NBR 14.653-2 para o maior grau de fundamentação. O coeficiente de correlação foi de 0,8568, indicando uma forte correlação entre as variáveis independentes e a variável dependente.

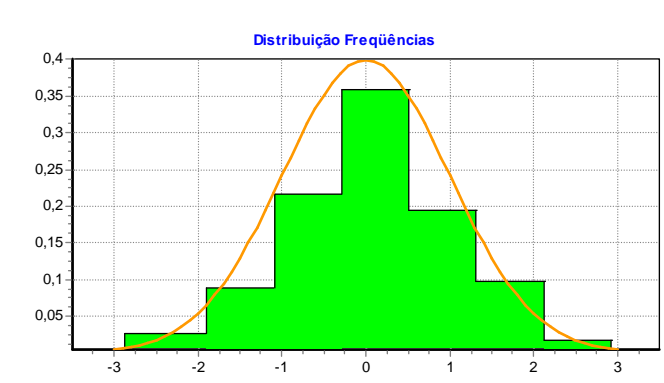


GRAFICO 4.24 – Distribuição de Frequência dos resíduos do MCRL

Na análise microeconômica do mercado de habitação de Belo Horizonte, o modelo inferido permitiu extrair informações importantes sobre a variação de preços, para o período estudado. Este indica uma valorização média no período estudado (20 meses) de 10,97%. Outra informação muito útil é com relação à localização dos imóveis. Para um apartamento com características médias, o preço

do imóvel varia aproximadamente de 700,00 R\$/m² a 1.050,00 R\$/m², ou seja, a localização pode representar até 50% do valor do imóvel.

O modelo apresentou um bom poder de predição, apesar de possuir uma tendência de subavaliar imóveis de valores mais altos, que pode ser verificado no GRÁFICO 4.25, que apresenta os preços versus valores estimados do modelo. A reta na diagonal deste gráfico representa a situação em que os valores estimados para os imóveis são exatamente iguais aos preços praticados. A segunda reta é uma regressão linear simples entre preços e valores estimados e tem por objetivo a visualização do grau de ajuste dos preços e valores estimados.

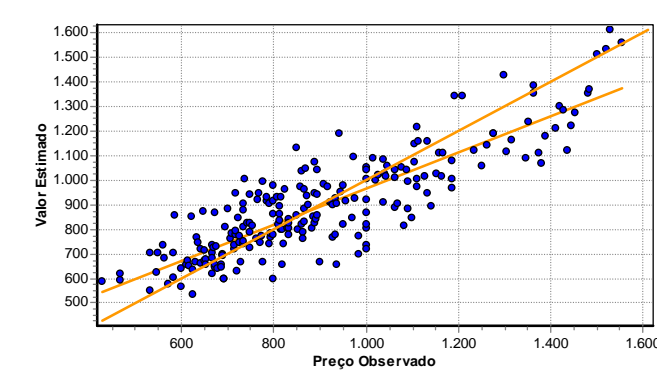


GRAFICO 4.25 – Valores estimados por Preços observados

O modelo empírico encontra-se explicitado na equação 4.2 a seguir:

$$\begin{aligned} \text{Valor Unitário} = & e^{(+7,039344981+0,0008226328371 * \text{Setor Urbano}+0,1613379329 * \\ & \ln(\text{Nível/Elevador})+0,005733353163 * \text{Data do Evento}+0,001348528869 * \\ & \text{Total Vagas}-0,5026981916 * \ln(\text{Área Coberta})+0,002915663288 * \text{Área Desc.}+ \\ & 0,1168737474 * \text{Dormitórios} + 0,009908174092 * \text{Equipamentos}- \\ & 0,05386150912 * \ln(\text{Unidades por Acesso})+0,001020533186 * \text{Padrão} + \\ & 0,0009130749008 * \text{Conservação}) \end{aligned} \quad (4.2)$$

Os resultados obtidos com o MCRL via MQO apresenta resultados estatísticos aceitáveis, mas não considerou a possibilidade da existência de autocorrelação espacial, o que pode invalidar este modelo, tendo em vista a violação do pressuposto básico de que todas as variáveis importantes foram incluídas na modelagem.

4.6.2 REGRESSÃO ESPACIAL

Ao modelo tradicional de preços hedônicos, que considerada o preço da habitação (P) como função de seus atributos, notadamente aqueles referentes à localização (L), estruturais (E) e temporais (T), será acrescida a dimensão espacial. A rigor, antes de proceder aos testes de inferência espacial, não se pode concluir que o modelo MCRL, estimado na seção anterior, seja eficiente ou o mais adequado, apesar dos bons resultados estatísticos observados. A nova forma do modelo fica definida como na equação (4.3),

$$P = f(L, E, T, \beta, WP) + \varepsilon \quad (4.3)$$

onde f é indicativo da forma funcional, β são os parâmetros do modelo, W a matriz de pesos espacial e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados e transformações realizadas no modelo MCRL da seção anterior.

Para a estimação empírica do modelo (4.3) será utilizada a Regressão Espacial, via Máxima Verossimilhança. Nesta abordagem será investigada a existência da autocorrelação espacial dos dados.

Para a determinação dos parâmetros necessários ao cálculo da regressão espacial, foi utilizado o programa EasyKrig, para a construção do semi-variograma e determinação da distância máxima de influência do preço de um imóvel em relação a seus vizinhos. A FIGURA 4.3 abaixo indica a posição dos dados no espaço urbano, considerando a latitude e longitude em coordenadas UTM, mostrando a região dos bairros onde foram coletados preços acima de 2.000 R\$/m². Os imóveis com preços unitários superiores a este valor foram eliminados da modelagem,

conforme indicado na seção 4.5. Os demais dados foram utilizados para a construção da matriz de pesos espaciais.

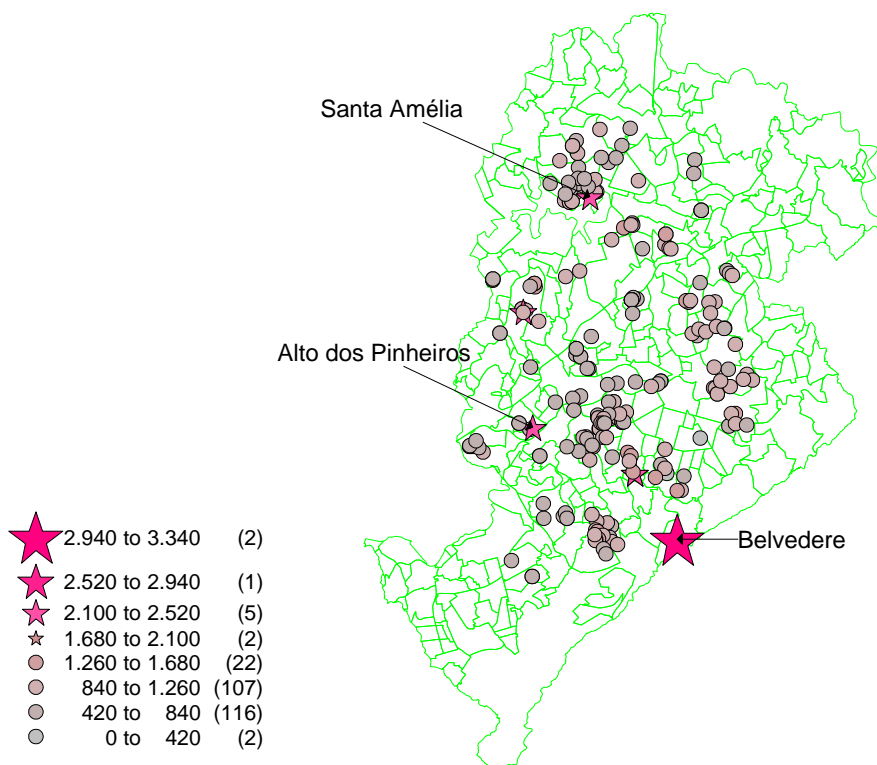


FIGURA 4.3 – Dados de mercado (latitude x longitude)

Em seguida, foi construído o semi-variograma, cujos resultados podem ser visualizados no GRÁFICO 4.26,

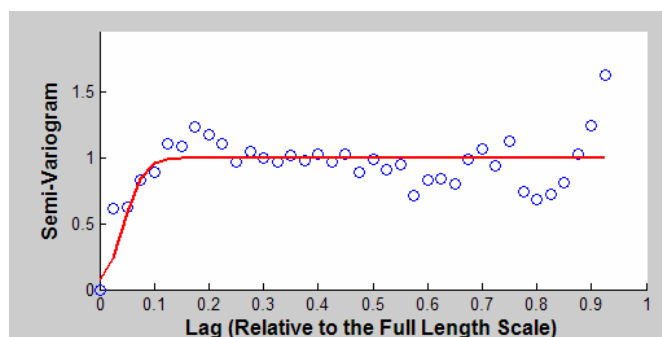


GRAFICO 4.26 – Semi-Variograma para os dados de mercado de Belo Horizonte / MG

onde estima-se a influencia de um evento de mercado em relação a sua vizinhança em um raio máximo de até 2,8 km (aproximadamente 0,2 da distancia

máxima coletada, que foi de 14 km). Este será o limite das distâncias para a construção da matriz de vizinhança W .

O diagnóstico da dependência espacial na amostra foi realizado utilizando-se os testes LM Robusto (erro) e LM Robusto (defasagem) sobre os resíduos do modelo estimado na tabela 4.4. Trabalhou-se com uma matriz de vizinhança W normalizada por linha. A tabela 4.5 apresenta os resultados para o diagnóstico realizado. O teste LM Robusto (defasagem) possui valor absoluto maior que o teste LM Robusto (erro), ou seja, o efeito vizinhança parece estar afetando os preços de habitação na cidade de Belo Horizonte / MG.

TABELA 4.5 Diagnóstico da Dependência Espacial

Teste MI/GL	Valor	Probabilidade
LM Robusto (erro)	5,73	0,01665
LM Robusto (defasagem)	16,7544	0,00004

Diante dessa possibilidade de autocorrelação espacial entre os preços de habitação ou efeito vizinhança, a seguir, estima-se o Modelo de Defasagem Espacial de Preços. Então, foi realizada uma regressão do preço de oferta ou transação sobre as variáveis independentes relacionadas na TABELA 4.3 utilizando o modelo funcional (4.3). Foi utilizado o software SisPlanV para o cálculo da regressão espacial, incluindo a variável de defasagem espacial.

Pelos resultados obtidos na TABELA 4.6 e no Gráfico 4.28, verifica-se que os testes de hipóteses (estatística t) e as hipóteses básicas da normalidade dos erros aleatórios foram inicialmente aceitos, considerando o modelo como consistente para a avaliação de apartamentos no mercado de Belo Horizonte.

TABELA 4.6 – Resultados estatísticos do modelo de preços hedônicos com a Regressão Espacial

Variáveis	Transformação	b (regressor)	t - Student	α (%) (bi caudal)
Setor Urbano	X	0,000706624	7,2	< 0,01
Nível/Elevador	Ln(X)	0,178771353	4,10	0,05
Data do Evento	X	0,005513866	2,40	1,95
Total Vagas	X	0,001296873	6,83	< 0,01
Área Coberta	Ln(X)	-0,506194192	-9,33	< 0,01
Área Descoberta	X	0,002978795	4,80	< 0,01
Dormitórios	X	0,120864298	5,17	< 0,01
Equipamentos	X	0,009361433	5,33	< 0,01
Unidades p/ Acesso	Ln(X)	-0,05063036	-3,97	< 0,01
Padrão Acabamento	X	0,000956406	6,37	< 0,01
Estado Conservação	X	0,00082779	5,74	< 0,01
WLn(Y)	X	0,262942164	3,696	0,02
Preços Unitários	Ln(Y)	+5,343664799		

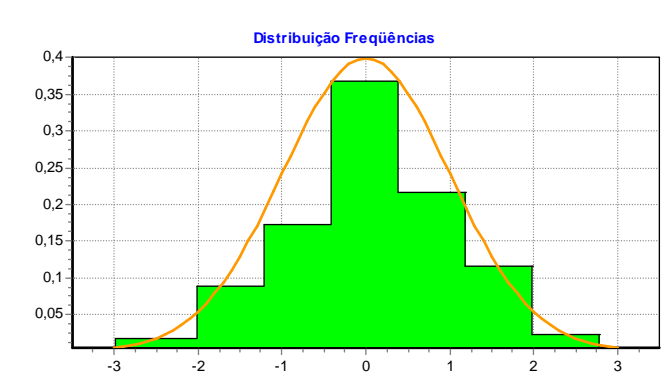


GRAFICO 4.28 – Distribuição de Frequência dos resíduos do MCRL

O teste de hipótese, usando a estatística t, para todas as variáveis independentes, apresentou significância abaixo de 10%, e o f calculado (análise de variância) foi de 56,25 com a significância do modelo abaixo de 0,01%, limites recomendados pela NBR 14.653-2 para o maior grau de fundamentação. O coeficiente de correlação foi de 0,8646, superior ao obtido pelo MCRL e indicando uma forte correlação entre as variáveis independentes e a variável dependente.

Assim, pode-se concluir que os preços dos imóveis não dependem somente das variáveis explicativas utilizadas pelo MCRL, mas também são fortemente influenciados, de maneira positiva, pelos preços que ocorrem na vizinhança.

O modelo inferido permitir novamente extrair informações importantes sobre a variação de preços no mercado de Belo Horizonte, no período de estudo. Este indica uma valorização média no período estudado para os imóveis em 11,11%. Outra informação muito útil é com relação à localização dos imóveis. Para um apartamento com características médias, o preço do imóvel varia aproximadamente de 850,00 R\$/m² a 1.200,00 R\$/m², ou seja, a localização pode representar até 40% do valor do imóvel.

O modelo apresentou um bom poder de predição, melhor do que o estimado pelo MCRL e que pode ser verificado no GRÁFICO 4.29, que apresenta os preços por valores estimados do modelo.

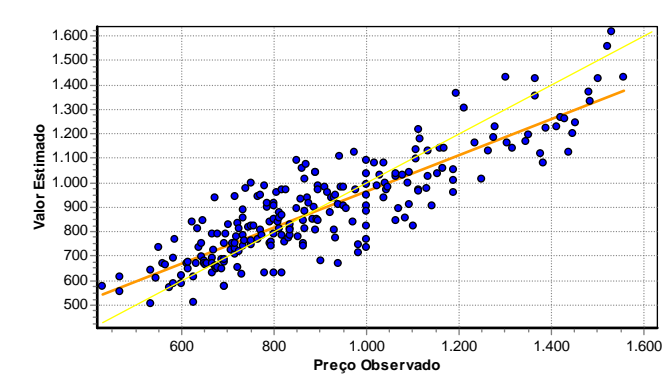


GRAFICO 4.29 – Valores estimados por Preços observados

Os resultados obtidos com a Regressão Espacial foram satisfatórios, tendo sido incorporada a microlocalização dos imóveis através da inclusão da variável de defasagem espacial. Os resultados da estimação espacial apresentados na tabela 4.6 indicam que o coeficiente ρ , da variável defasada espacialmente $W \ln(Y)$ é positivo e estatisticamente significativa a menos de 1%, implicando na existência de um forte efeito de defasagem espacial nos dados da amostra coletada. Neste caso, os preços dos imóveis não mais dependem das variáveis estruturais, de localização e temporais, mas também é fortemente influenciado pelos preços de seus vizinhos até a uma distância de 2,8 km.

O modelo empírico encontra-se explicitado na equação (4.4) a seguir:

$$\begin{aligned} \text{Valor Unitário} = e^{(+5,343664799 + 0,0007066240486 * \text{Setor Urbano} +} \\ 0,1787713532 * \ln(\text{Nível/Elevador}) + 0,00551386571 * \text{Data} +} \\ 0,00129687329 * \text{Total Vagas} - 0,5061941919 * \ln(\text{Área Cob}) +} \\ 0,002978794592 * \text{Área Desc.} + 0,1208642976 * \text{Dormitórios} +} \\ 0,009361433193 * \text{Equipamentos} - 0,05063036011 * \ln(\text{Unidades por Acesso}) +} \\ 0,0009564062885 * \text{Padrão Acabamento} +} \\ 0,0008277902485 * \text{Estado Conservação} + 0,2629421637 * w * y) \end{aligned} \quad (4.4)$$

4.6.3 REDES NEURAIS ARTIFICIAIS

O modelo tradicional de preços hedônicos, que considerada o preço da habitação (P) com função de seus atributos, notadamente aqueles referentes à localização (L), estruturais (E) e temporais (T), será modelado a partir da utilização das RNAs. Neste momento não será acrescida a dimensão espacial. A forma do modelo fica definida como na equação (4.5),

$$P = f(L, E, T, w) + \varepsilon \quad (4.5)$$

onde f é indicativo da forma funcional, w os pesos a serem ajustados durante a fase de treinamento da rede e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados, porém sem transformações, do modelo MCRL da seção 4.6.1.

Para a estimação empírica do modelo (4.5) será utilizada uma arquitetura de RNA, com aprendizado via algoritmo de treinamento *Levenberg-Marquadt*. A estrutura da rede será empiricamente composta por 6⁴⁶ (seis) neurônios na camada escondida e a função de transferência será a tangente hiperbólica, para a camada intermediária, e linear para a camada de saída. Foram selecionadas 15 épocas para o treinamento e o erro quadrático mínimo para a finalização do treinamento foi de

⁴⁶ Definido inicialmente de forma empírica, sem a utilização da validação cruzada.

0,00001. O conjunto de dados para o treinamento foi normalizado no intervalo de 0,2 a 0,8.

Com base nestes parâmetros, foi iniciado o treinamento da rede considerando como variável de saída o preço de oferta ou transação e as variáveis de entrada as relacionadas na TABELA 4.3. Nesta etapa do trabalho não foram introduzidos os métodos para controle e supervisão do treinamento da RNA, que serão incluídos nas próximas seções.

O tempo total para a fase de treinamento foi de aproximadamente 3 segundos, em um computador *pentium* 4 com 512MB de memória *ram*. A variância obtida nos resultados da RNA foi bem grande, e foram necessários mais de 100 treinamentos para a estabilização e escolha do modelo neural a ser utilizado.

Os resultados obtidos com a modelagem estão listados na tabela 4.7:

TABELA 4.7 – Resultados da estimação com as RNAs

Variáveis	Valor Médio	Variação (%)
Setor Urbano	484,3142	3,98
Nível/Elevador	8,0575	3,56
Data do Evento	5,3407	0,99
Total Vagas	115,0442	4,04
Área Coberta	80,5880	-8,18
Área Descoberta	5,4513	5,62
Dormitórios	2,7566	2,04
Equipamentos	8,4735	2,65
Unidades p/ Acesso	11,4513	-1,80
Padrão Acabamento	250,7080	4,70
Estado Conservação	473,8938	2,88
Preço Unitário Estimado	876,88	

A coluna “Variação (%)” da tabela 4.7 acima indica a variação no valor estimado dos imóveis pela RNA quando a variável em estudo apresenta um acréscimo de 10% de sua amplitude. O modelo inferido permitiu novamente extrair informações importantes sobre a variação de preços no mercado de Belo Horizonte, no período de estudo. Este indica uma valorização média dos imóveis em aproximadamente 10% no período estudado. Outra informação muito útil é com

relação á localização dos imóveis. Para um apartamento com características médias, o preço do imóvel varia aproximadamente de 700,00 R\$/m² a 1.050,00 R\$/m², ou seja, a localização pode representar até 50% do valor do imóvel, mesmo valor encontrado pelo MCRL.

O coeficiente de correlação foi de 0,8960, superior aos demais modelos, o que indica uma forte correlação entre as variáveis de entrada e a variável de saída e a distribuição dos resíduos sinaliza que o modelo possui tendências à normalidade, conforme observa-se no GRÁFICO 4.30,

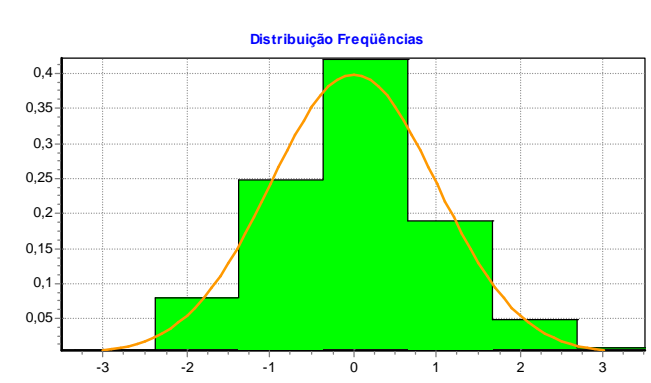


GRAFICO 4.30 – Distribuição de Freqüência dos resíduos de RNAs

O modelo apresentou um bom poder de predição, melhor do que o modelo da Regressão Espacial e da Regressão Linear, que pode ser verificado no GRÁFICO 4.31, que apresenta os preços observados por valores estimados do modelo.

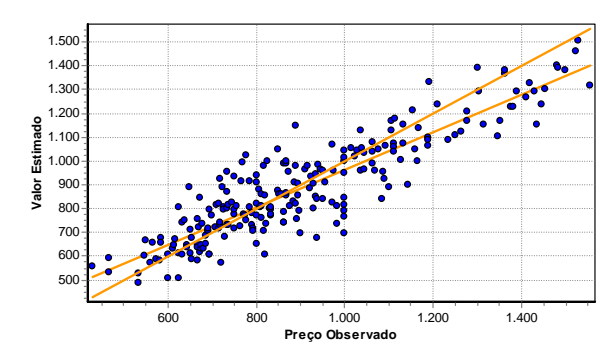


GRAFICO 4.31 – Valores estimados por Preços observados

Os resultados obtidos com a RNA foram mais satisfatórios do que os obtidos com os demais modelos, apesar de não ter sido considerada a existência de autocorrelação espacial. A aparente superioridade do modelo de RNA provavelmente se deve ao seu alto poder de mapeamento de funções não lineares, principalmente considerando que os dados coletados apresentaram grande heterogeneidade. Contudo, existe a necessidade de se proceder à validação dos resultados, que será tratada na próxima seção. Outro fato importante foi a dificuldade na seleção do modelo neural, sem a utilização de um processo de controle durante o treinamento. Considerando que a variabilidade dos resultados foi bem grande, a seleção do modelo se deu de forma empírica, através da análise do erro médio quadrático e das projeções elaboradas conforme a TABELA 4.7. Ficou evidenciada a necessidade da utilização de técnicas mais avançadas, como os procedimentos de poda e de *bagging*, descritos de forma resumida nas seções 3.4.10 e 3.4.11. A seguir o procedimento desta seção será repetido, com a inclusão da variável de defasagem espacial.

4.6.4 REDES NEURAS ARTIFICIAIS ESPACIAIS

Ao modelo tradicional de preços hedônicos, que considerada o preço da habitação (P) com função de seus atributos, notadamente aqueles referentes à localização (L), estruturais (E) e temporais (T), será incluída a variável de defasagem espacial WP . Assim, os dados serão modelados a partir da utilização das RNAs. Neste momento será acrescida a dimensão espacial, sem incluir o processo de poda ou *bagging*. A forma do modelo fica definida como na equação 4.6,

$$P = f(L, E, T, w, WP) + \varepsilon \quad (4.6)$$

onde f é indicativo da forma funcional, w os pesos a serem ajustados durante a fase de treinamento da rede, W a matriz de proximidade espacial e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados, porém sem transformações, do modelo MCRL da seção 4.6.1, incluindo a variável de defasagem espacial.

A arquitetura da rede está representada na FIGURA 4.4 abaixo:

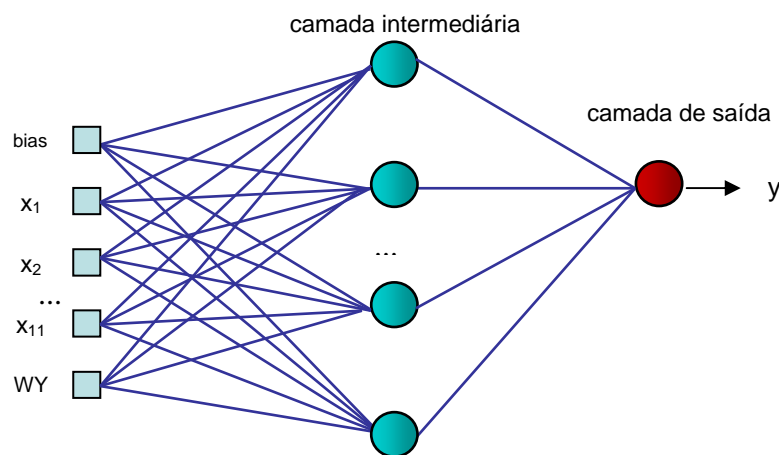


FIGURA 4.4 – RNA com a inclusão da variável de defasagem espacial – WP

Para a estimação empírica do modelo (4.6) será utilizada a RNA, via algoritmo de treinamento *Levenberg-Marquadt*. A estrutura da rede será composta por 6⁴⁷ (seis) neurônios na camada escondida e a função de transferência será a tangente hiperbólica, para a camada intermediária, e linear para a camada de saída. O número de épocas para o treinamento foi de 15 e o erro quadrático mínimo para a finalização do treinamento foi de 0,00001. Com base nestes parâmetros, foi iniciado o treinamento da rede considerando como variável de saída o preço de oferta ou transação e as variáveis de entrada as relacionadas na TABELA 4.3 (incluindo a variável Wy).

O tempo total para a fase de treinamento foi de aproximadamente 3 segundos, em um computador *pentium 4* com 512MB de memória *ram*.

⁴⁷ Definido inicialmente de forma empírica.

Os resultados obtidos com a modelagem estão listados na tabela 4.8:

TABELA 4.8 – Resultados da estimação com as RNAs Espaciais

Variáveis	Valor Médio	Variação (%)
Setor Urbano	484,3142	3,21
Nível/Elevador	8,0575	3,39
Data do Evento	5,3407	1,36
Total Vagas	115,0442	3,79
Área Coberta	80,5880	-6,97
Área Descoberta	5,4513	4,23
Dormitórios	2,7566	1,96
Equipamentos	8,4735	2,37
Unidades p/ Acesso	11,4513	-1,41
Padrão Acabamento	250,7080	3,78
Estado Conservação	473,8938	2,38
WP	904,8713	2,34
Preço Unitário Estimado	910,36	

O modelo inferido permitir novamente extrair informações importantes sobre a variação de preços no mercado de Belo Horizonte, no período de estudo. Este indica uma valorização média mensal dos imóveis em aproximadamente 15% no período estudado.

Outra informação muito útil é com relação à localização dos imóveis. Para um apartamento com características médias, o preço do imóvel varia aproximadamente de 700,00 R\$/m² a 1.000,00 R\$/m², ou seja, a localização pode representar até aproximadamente 43% do valor do imóvel, resultado próximo ao encontrado pela Regressão Espacial.

O coeficiente de correlação foi de 0,8780, inferior ao das RNAs sem a variável de defasagem, mas que indica uma forte correlação entre as variáveis de entrada e a variável de saída. A diminuição do C_c deste modelo, quando comparado com o modelo de RNA sem a variável de defasagem espacial, pode ser explicada pela inclusão de mais uma variável no modelo, sem alterar a estrutura da RNA utilizada.

A distribuição dos resíduos sinaliza que o modelo possui tendências à normalidade, conforme se observa no GRÁFICO 4.32,

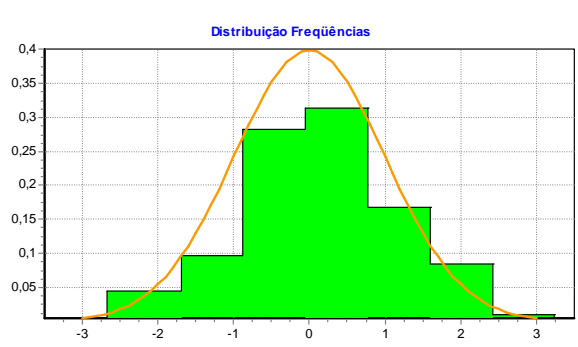


GRAFICO 4.32 – Distribuição de Frequência dos resíduos de RNAs

O modelo apresentou um bom poder de predição, melhor do que da Regressão Espacial e da Regressão Linear, que pode ser verificado no GRÁFICO 4.33 que apresenta os preços observados por valores estimados do modelo.

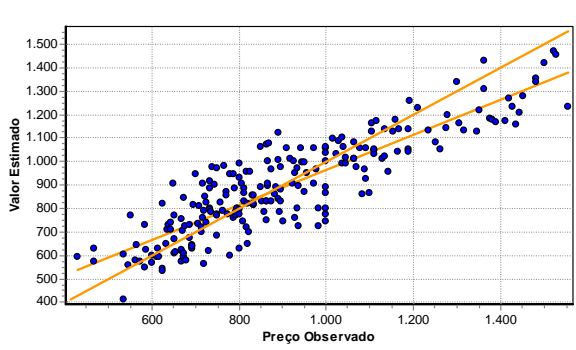


GRAFICO 4.33 – Valores estimados por Preços observados

Os resultados obtidos com as RNAs Espaciais foram mais satisfatórios do que os modelos de regressão, mesmo quando incluída a variável para tratamento da autocorrelação espacial. A aparente superioridade do modelo de RNAs provavelmente se deve ao seu alto poder de mapeamento de funções não lineares, principalmente considerando que os dados coletados apresentaram grande heterogeneidade. Contudo, existe a necessidade de se proceder à validação dos resultados, que será tratada na próxima seção.

4.6.5 VALIDAÇÃO DOS RESULTADOS

Considerando que foi elaborada apenas uma modelagem para a comparação entre os diversos tratamentos efetuados, e como não foi introduzido nenhum processo de validação dos resultados, esta seção irá tratar da validação cruzada e das técnicas mais utilizadas para o controle do processo de treinamento das RNAs.

Para realizar a validação cruzada, com o objetivo de escolher o melhor modelo para as avaliações no mercado habitacional para apartamentos, de Belo Horizonte, foram selecionadas então cinco amostras, cada uma com um conjunto de dados diferente para a validação e para a modelagem de dados. A amostra inicialmente coletada, composta de 255 dados de mercado, foi subdividida em cinco amostras, cada uma com dois conjuntos de dados, sendo um para treinamento e outro para validação. Para realizar este procedimento, a seleção dos dados de validação foi feita aleatoriamente, sendo repetida por 5 vezes, resultando em 5 amostras de dados, cada uma composta de um conjunto de dados para a modelagem e outro para a validação cruzada. Para o sorteio, 20% dos dados foram separados da modelagem para a validação dos resultados.

Em resumo, dos 255 dados coletados no período, 231 foram efetivamente utilizados (foram desconsiderados dados atípicos e outliers). Destes, 189 foram utilizados para a modelagem e 42 para a validação. Este procedimento foi repetido por 5 vezes, selecionando aleatoriamente os dados para modelagem e validação, mas mantendo a proporção de 80% e 20%.

A modelagem de dados, utilizando as RNAs, foi realizada com o uso do software SisReN – Sistema de Regressão Linear e RNAs e para as técnicas de poda

foi utilizado o pacote para Matlab – NNSYSID20. Para cada uma das cinco amostras foram realizadas sete modelagens com abordagens diferenciadas:

- a 1º utilizando a regressão espacial;
- a 2º com o MCRL;
- a 3º utilizando as RNAs, com as mesmas variáveis do MCRL, sem aplicar as técnicas de poda, mas utilizando a validação cruzada;
- a 4º utilizando as RNAs, com as mesmas variáveis do MCRL, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE) e a validação cruzada;
- a 5º utilizando as RNAs, com as mesmas variáveis da regressão espacial, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE) e a validação cruzada;
- a 6º utilizando as RNAs, com as mesmas variáveis do MCRL, aplicando Bagging;
- a 7º utilizando as RNAs, com as mesmas variáveis da regressão espacial, aplicando *bagging*.

Os resultados das modelagens acima foram obtidos seguindo os mesmos passos descritos nesta seção, com o uso do MCRL, da Regressão Espacial e das RNAs. Entretanto, nesta fase foi utilizada a validação cruzada para maior controle durante a fase de treinamento, bem como foram aplicadas as técnicas de PODA e de *bagging*.

A TABELA 4.09 contém os dados selecionados aleatoriamente para compor o conjunto de validação. Os demais dados foram utilizados para a modelagem.

TABELA 4.09 – Dados selecionados para a validação cruzada

Dados selecionados para a validação				
Amostra 1	Amostra 2	Amostra 3	Amostra 4	Amostra 5
5	4	4	3	6
10	6	13	4	11
13	16	16	15	13
15	27	17	17	22
23	30	18	27	27
28	33	30	61	35
55	34	34	81	52
57	50	51	82	53
71	51	54	84	66
74	55	55	87	75
75	66	65	88	76
86	69	67	92	77
89	79	72	98	85
96	80	75	108	86
100	82	78	114	92
105	84	80	117	95
106	86	93	118	104
107	92	95	123	114
112	95	96	127	118
124	103	99	130	128
125	110	102	132	134
141	111	107	137	135
145	131	111	138	139
151	132	115	145	149
154	134	121	152	152
156	154	130	155	154
165	158	133	168	163
168	171	138	172	175
180	180	153	192	183
190	188	158	204	187
197	192	180	207	193
203	212	183	211	194
210	216	191	212	209
219	229	194	214	215
227	239	196	215	217
228	240	199	222	232
230	243	204	226	236
236	248	208	241	248
243	249	221	242	250
246	250	223	248	251
250	251	229	252	254
254	252	246	253	255

Em seguida será elaborada uma análise dos resultados obtidos.

4.7 ANÁLISE DOS RESULTADOS DAS MODELAGENS

Na tabela 4.10 foram listados os resíduos quadráticos médios na validação dos resultados, conforme 4.7,

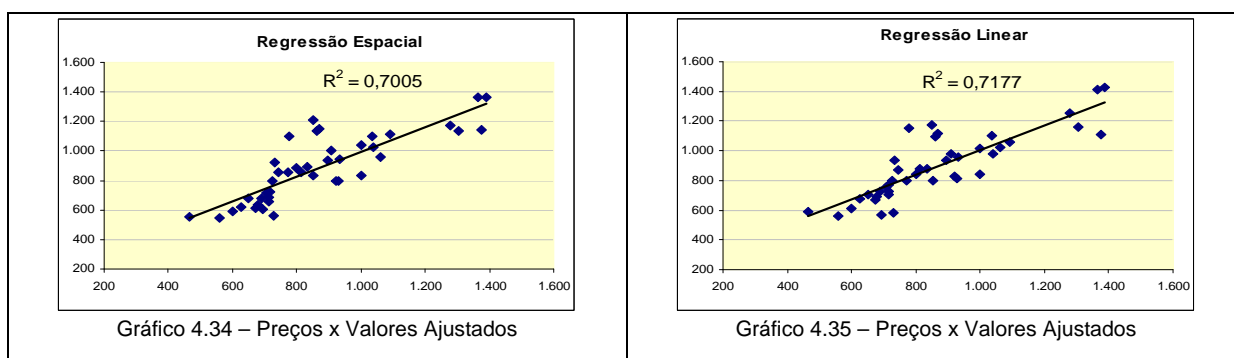
$$\varepsilon = \frac{\sum(P-V)^2}{n} \quad (4.7)$$

onde ε é igual ao resíduo quadrático médio na validação dos modelos, P os preços dos imóveis que compõe o conjunto de validação e V os valores estimados para estes mesmos dados.

TABELA 4.10 – Resíduos quadráticos médios do processo de validação

	Resíduo Quadrático Médio para a validação				
Modelos	1	2	3	4	5
Regressão Espacial	16.578	24.878	22.580	16.731	17.303
Regressão Linear	15.713	28.099	26.370	19.764	19.371
RNA	13.565	23.532	20.343	15.488	18.475
RNA Prune	13.394	15.622	17.173	15.627	14.215
RNA Wy Prune	15.785	15.662	17.078	17.052	15.157
RNA bagging		15.424			
RNA Wy bagging		15.771			

Para melhor visualizar os resultados, os gráficos de 4.34 a 4.58 apresentam os Preços x Valores Ajustados para cada uma das amostras.



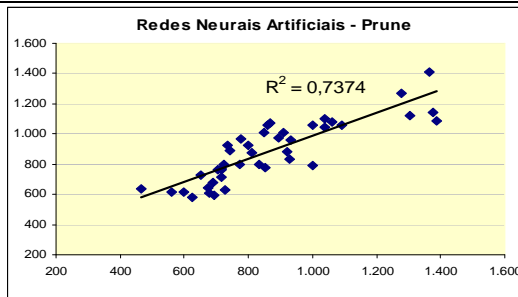


Gráfico 4.36 – Preços x Valores Ajustados

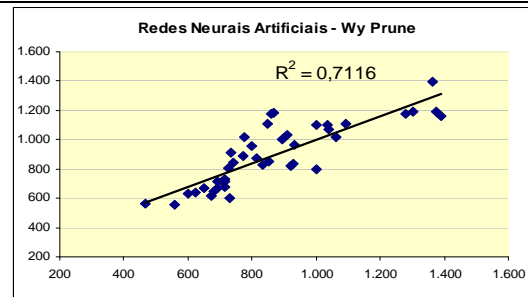


Gráfico 4.37 – Preços x Valores Ajustados

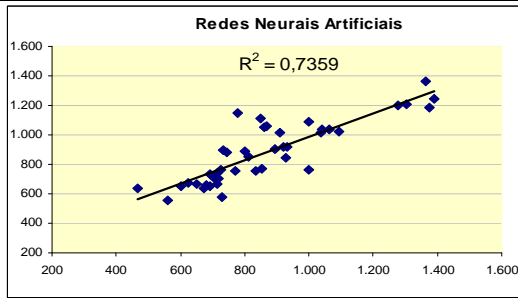


Gráfico 4.38 – Preços x Valores Ajustados

Os gráficos de números 4.34 a 4.38 representam os resultados para a amostra 1.

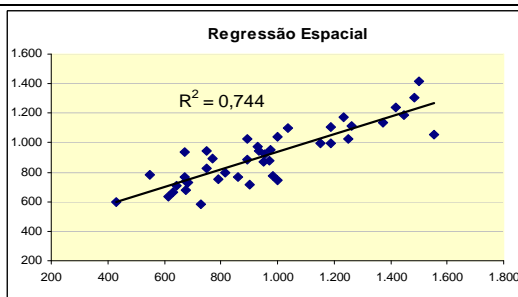


Gráfico 4.39 – Preços x Valores Ajustados

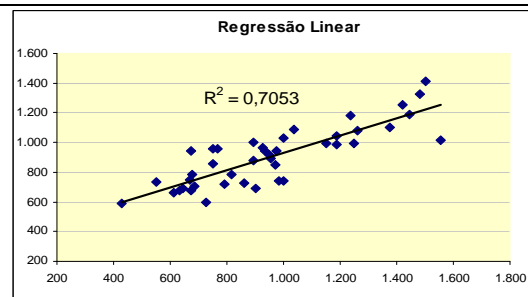


Gráfico 4.40 – Preços x Valores Ajustados

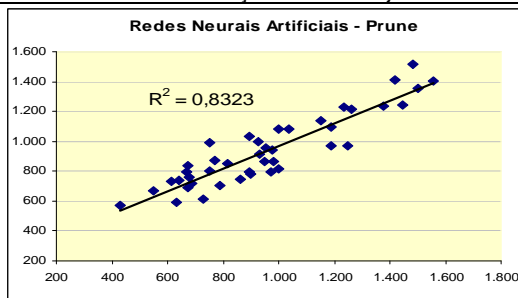


Gráfico 4.41 – Preços x Valores Ajustados

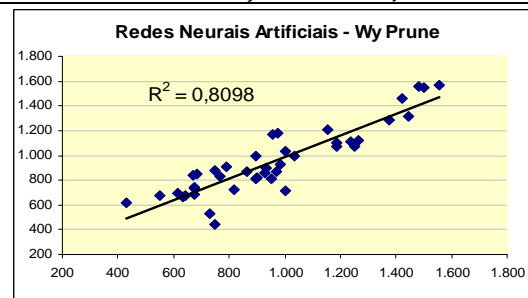


Gráfico 4.42 – Preços x Valores Ajustados

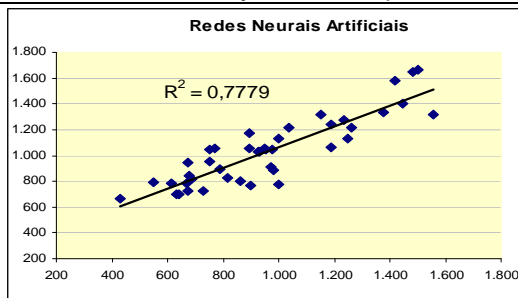


Gráfico 4.43 – Preços x Valores Ajustados

Os gráficos de números 4.39 a 4.43 representam os resultados para a amostra 2.

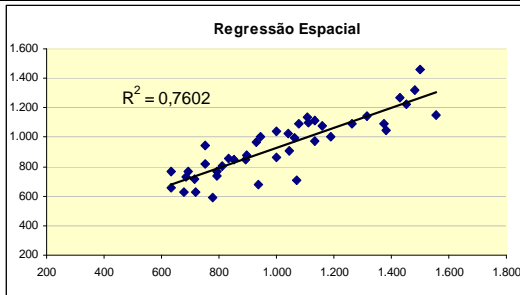


Gráfico 4.44 – Preços x Valores Ajustados

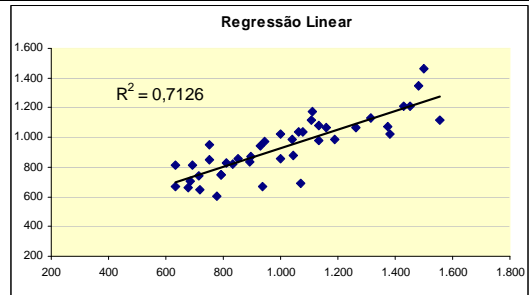


Gráfico 4.45 – Preços x Valores Ajustados

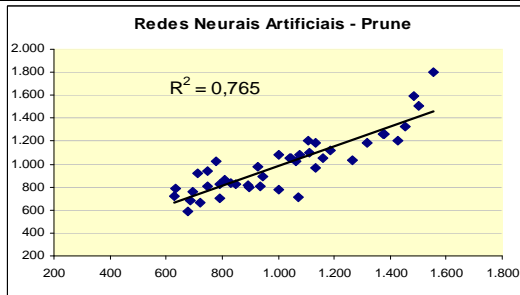


Gráfico 4.46 – Preços x Valores Ajustados

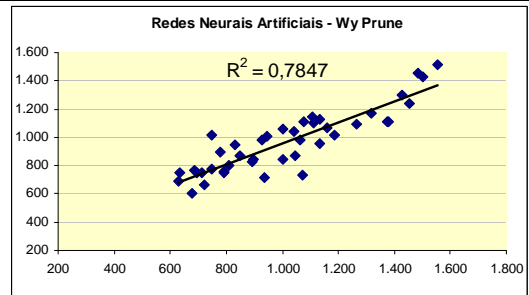


Gráfico 4.47 – Preços x Valores Ajustados

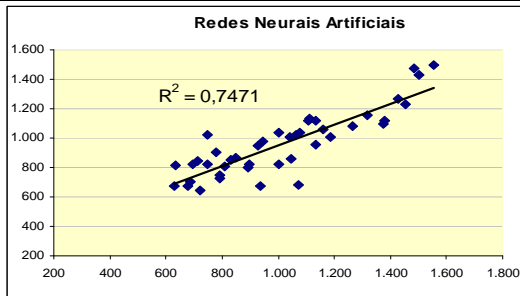


Gráfico 4.48 – Preços x Valores Ajustados

Os gráficos de números 4.44 a 4.48 representam os resultados para a amostra 3.

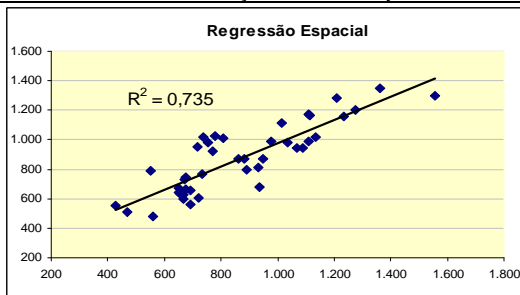


Gráfico 4.49 – Preços x Valores Ajustados

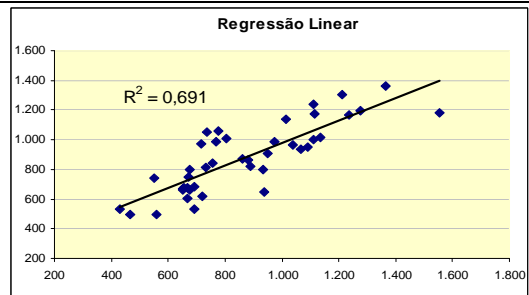


Gráfico 4.50 – Preços x Valores Ajustados

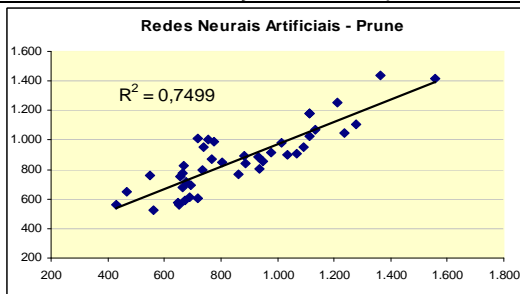


Gráfico 4.51 – Preços x Valores Ajustados

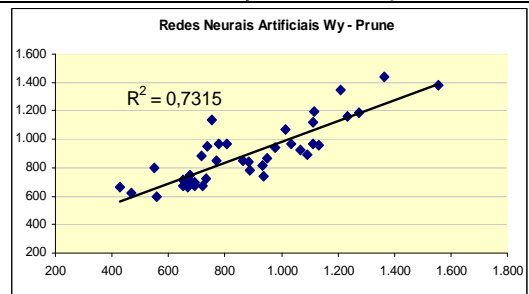


Gráfico 4.52 – Preços x Valores Ajustados

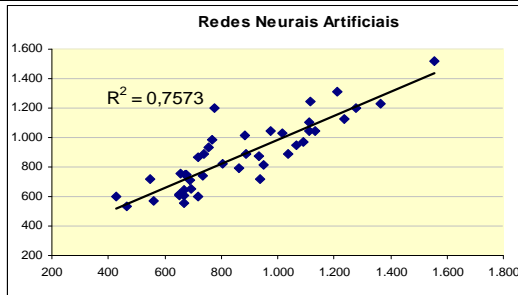


Gráfico 4.53 – Preços x Valores Ajustados

Os gráficos de números 4.49 a 4.53 representam os resultados para a amostra 4.

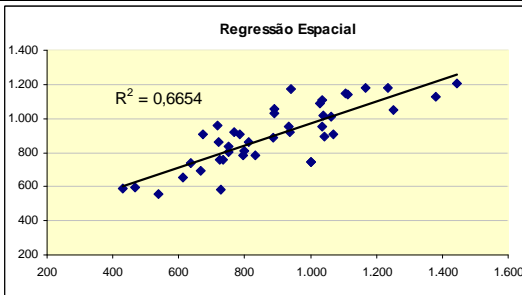


Gráfico 4.54 – Preços x Valores Ajustados

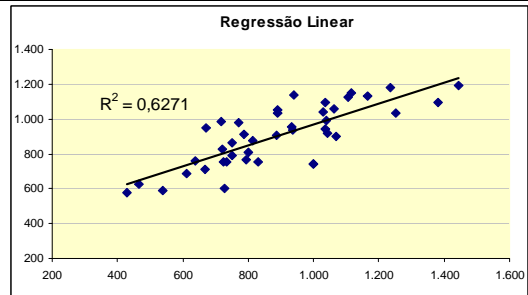


Gráfico 4.55 – Preços x Valores Ajustados

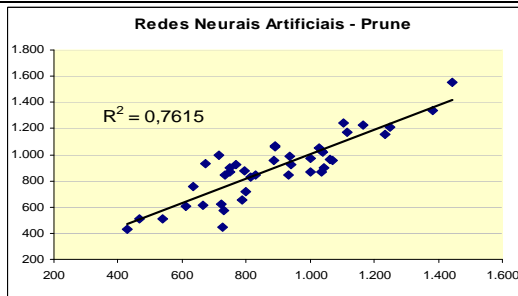


Gráfico 4.56 – Preços x Valores Ajustados

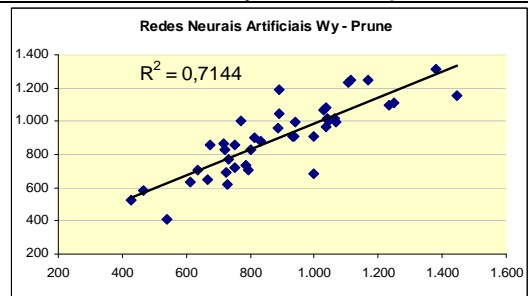


Gráfico 4.57 – Preços x Valores Ajustados

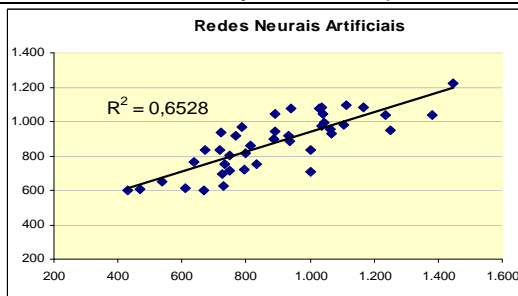


Gráfico 4.58 – Preços x Valores Ajustados

Os gráficos de números 4.54 a 4.58 representam os resultados para a amostra 5.

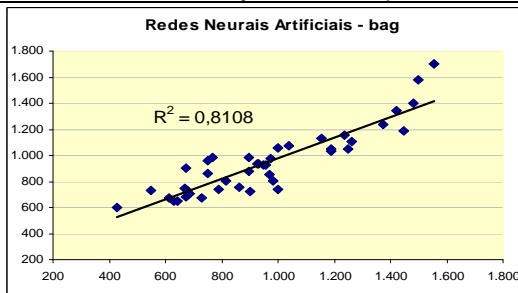


Gráfico 4.59 – Preços x Valores Ajustados

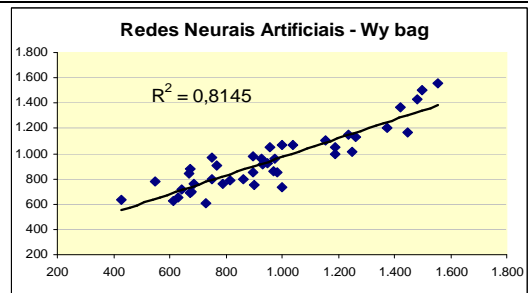


Gráfico 4.60 – Preços x Valores Ajustados

Os gráficos de números 4.59 e 4.60 representam os resultados para a amostra 2 com a utilização do *bagging*.

A reta existente nos gráficos acima é uma regressão linear simples entre o preço do imóvel e o valor estimado pelo modelo para os dados selecionados para validação. Esta reta tem como finalidade fornecer um indicativo do grau de correlação entre preço e valor para o conjunto de validação. No gráfico também está indicado o coeficiente de determinação R^2 calculado pela regressão. A TABELA 4.11 descreve os resultados obtidos com cada modelo, considerando o erro quadrático médio e o R^2 calculado, para cada amostra e tratamento efetuado.

TABELA 4.11 – Resultados do processo de validação

Modelo / Amostra	Erro Quadrático Médio					R^2 - Preços x Valores				
	1	2	3	4	5	1	2	3	4	5
Regressão Espacial	16.578	24.878	22.580	16.731	17.303	0,70	0,74	0,76	0,74	0,66
Regressão Linear	15.713	28.099	26.370	19.764	19.371	0,72	0,71	0,71	0,69	0,63
RNA	13.565	23.532	20.343	15.488	18.475	0,74	0,78	0,75	0,76	0,65
RNA Prune	13.394	15.622	17.173	15.627	14.215	0,74	0,83	0,77	0,75	0,76
RNA Wy Prune	15.785	15.662	17.078	17.052	15.157	0,71	0,81	0,79	0,73	0,71
RNA Bagging	15.339	15.424					0,81			
RNA Wy Bagging		15.771					0,81			

As colunas dos erros quadráticos médios permitem identificar, por amostra, o modelo que apresentou o menor erro na validação dos resultados. Na análise visual desta tabela fica evidenciado que em todos os modelos, os melhores resultados, em sua maioria, são os que utilizaram as RNAs e que o processo de poda possibilitou a obtenção de melhores resultados. Entretanto, esta coluna não é apropriada para a comparação entre os resultados obtidos com os diversos modelos para todas as amostras, pois estas diferem nas escalas das variáveis e os resultados com base nesta comparação são enganosos. Assim, as colunas referentes ao R^2 são melhores para a comparação direta entre todos os modelos e novamente os modelos de RNAs apresentaram um melhor ajuste ao conjunto de validação. Para a amostra 2, quatro modelos apresentaram coeficientes de correlação acima de 0,80 , indicando que as RNAs obtiveram um bom poder de

predição para novos imóveis, quando comparadas com o MCRL e a Regressão Espacial.

Dos quatro modelos com coeficiente de determinação superior a 0,80 na validação (amostra 2), foi selecionado o modelo de RNA *Wy bagging*, em função dos demais resultados obtidos, indicando ser este o modelo que melhor descreve o mercado de apartamentos da cidade de Belo Horizonte. O coeficiente de correlação C_c ⁴⁸ foi de 0,92 na fase de treinamento. O gráfico 4.60 indica o poder de predição deste modelo para o conjunto de dados da validação.

Na análise microeconômica do mercado habitacional de Belo Horizonte, com o modelo selecionado, obteve-se uma valorização dos preços dos imóveis, para o período estudado, de 16,77%. Para este mesmo período, o IPEAD – FACE / UFMG publicou o índice de 17,73%, bem próximo ao índice calculado pelo modelo. A macrolocalização dos imóveis representa, em média, 43% de seu valor, resultado este também compatível com os calculados pelos demais modelos que utilizaram a variável de defasagem espacial *Wy*. O modelo utilizando a RNA com *prunning* obteve um R^2 na validação superior aos demais modelos, porém nas demais análises⁴⁹, os resultados foram inferiores ao modelo selecionado.

⁴⁸ O coeficiente de correlação é a raiz quadrada do coeficiente de determinação e indica o grau de correlação entre o preço e o valor estimado pelo modelo.

⁴⁹ Análise dos resíduos, da aderência à distribuição normal e das variações percentuais na variável de saída, quando, para cada variável de entrada, é acrescido 10% da amplitude ao valor médio.

5. CONCLUSÕES

Pelos resultados encontrados neste trabalho, fica evidenciada a importância da utilização de novas metodologias de modelagem por RNAs e pela Regressão Espacial nos estudos dos fenômenos relacionados com o comportamento do mercado imobiliário. Na análise empírica realizada na cidade de Belo Horizonte / MG, verificaram-se que os modelos de RNAs tiveram um melhor ajuste tanto na fase de modelagem quanto na fase de validação cruzada, inclusive quando comparados com os modelos da Regressão Espacial. Também foram identificados fortes indícios de dependência espacial nos preços dos apartamentos, comprovando-se, desta forma, que o MCRL não é adequado para realização de estudos dessa natureza e que avaliações confiáveis, caracterizadas pela não tendenciosidade, eficiência e consistência, somente podem ser obtidas com a utilização das RNAs e, em segundo plano, pelos Modelos de Regressão Espacial.

Um dos problemas que têm limitado o uso das RNAs na Engenharia de Avaliações é explicitar o modelo matemático, ou seja, a dificuldade em demonstrar, de forma simplificada, como foi obtido o valor do imóvel, e não somente a visualização do desenho da rede neural, com a apresentação dos pesos das conexões e da função de transferência utilizada. Para as avaliações em massa, principalmente para fins tributários, este é um dos problemas a ser tratado pelo especialista.

As aplicações de RNAs na engenharia de avaliações não foram relatadas com sucesso, por diversos autores, citado em González⁵⁰ (2003), que relaciona algumas aplicações nesta área, com resultados não animadores. Contudo, diversos

⁵⁰ Aplicação de técnicas de descobrimento de conhecimento em base de dados e de inteligência artificial em avaliação de imóveis: págs. 193 a 196.

dos problemas citados já foram solucionados atualmente com a utilização de algoritmos eficientes e a utilização de um processo de controle para o treinamento e validação das RNAs, não sendo mais fatores impeditivos para a modelagem.

Neste trabalho foram obtidos melhores resultados através das RNAs, principalmente quando da aplicação das técnicas de *Prunning* e de *Bagging*, que proporcionaram um melhor poder de generalização às RNAs.

Os resultados obtidos com a Regressão Espacial foram também superiores aos obtidos com a regressão linear, contudo inferiores aos modelos de RNAs. A introdução da variável de defasagem espacial nos modelos de RNAs propiciou melhorias nos resultados, indicando que esta variável deve ser considerada nos modelos a serem empregados nas avaliações em massa. A separação da amostra em conjuntos de dados para modelagem e validação permite obter resultados consistentes na aplicação das RNAs, quando comparados com a Regressão Linear e a Regressão Espacial, originalmente proposta por Anselin(1998).

BIBLIOGRAFIA

- ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS: Norma Técnica de Avaliação de Bens, Procedimentos Gerais – NBR 14653:1, Rio de Janeiro: 2001.
- ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS: Norma Técnica de Avaliação de Bens, Imóveis Urbanos – NBR 14653:2, Rio de Janeiro: 2004.
- ANSELIN, L., “Spatial Econometrics: Methods and Models”. Dordrecht: Kluwer Academic, 1988.
- ANSELIN, L., “Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity”, *Geographical Analysis* 20, 1-17, 1988a.
- ANSELIN, L., “SpaceStat: A program for the Statistical Analysis of Spatial Data” (Departament of Geography and Departament of Economics, University of California, Santa Barbara, CA), 1990.
- ANSELIN, L., Rey, S., “Properties of tests for spatial dependence in linear regression models”, *Geographic Analysis*, 23, 112-31, 1991.
- ANSELIN, L., Bera, A., “Spatial dependence in linear regression models with an introduction to spatial econometrics”, *Handbook of Applied Economic Statistics*, Giles: Marcel Dekker, 1991.
- ANSELIN, L., “Spatial Econometrics”, Bruton Center, School of Social Sciences, University of Texas, Dallas, Richardson, 1999.
- BRAGA, A. P., CARVALHO, A. P. L. F., LUDEMIR, T. B., “Redes Neurais Artificiais: Teoria e Aplicações”, Rio de Janeiro: LTC, 2000.
- BRAGA, A. P., CARVALHO, A. P. L. F., LUDEMIR, T. B., “Sistemas Inteligentes – Fundamentos e Aplicações”, RECOPE-IA – Rede Cooperativa de Pesquisa em Inteligência Artificial, Manole, págs. 141 a 168, 2003.
- BREIMAN, L., “Bagging Predictors”, Technical Report No. 421, 1994.
- BUSSAB, W. O., MORETTIN, P. A., *Estatística Básica*, 5ª Ed., São Paulo, Editora SARAIVA, 2004.

- CHARNES, A., W. W. Cooper and Rhodes, Measuring the Efficiency of Decision – Making Units, *European Journal of Operational Research*, 2, No.6 - 1978.
- COSTA NETO, P. L. O., *Estatística - 18º Reimpressão*, São Paulo: Editora EDGARD BLUCHER, 2000.
- CUN, Y. Le, DENKER, J. S., SOLLA, S. A., *Optimal Brain Damage: Advances in Neural Information Processing Systems*, vol.2, Morgan Kaufmann, San Mateo, p 598-605, 1989.
- DANTAS, R. A., Sales, F., CHAVES, R., MAGALHÃES, A., “Uma nova metodologia para avaliação de imóveis utilizando regressão espacial,” *Anais do XI Congresso Brasileiro de Engenharia de Avaliações e Perícias*, Vitória/ES, 2001.
- DANTAS, R. A, ROCHA, F. S. e MAGALHÃES, A., “La importancia de la Regresión Espacial en la Tasación de Inmuebles”, *Anais do I Congreso Internacional en Tasación Y Valoración*, Valencia, Espanha, 2002a.
- DANTAS, R. A, ROCHA, F. S. e MAGALHÃES, A., “Modelos Espaciales Aplicados en la Tasación de Inmuebles”, *Anais do XX Congreso Panamericano de Valuación*, Buenos Aires, Argentina, 2002b.
- DANTAS, R. A., *Engenharia de Avaliações – Uma Introdução à Metodologia Científica – Ed. PINI - São Paulo*, 2001.
- FAHLMAN, S. E., Libiere, C., *Advances in Neural Information Processing Systems*, Vol. 2, D. S. Touretzsky Ed., Morgan Kaufmann, San Mateo, p 524, 1990.
- GONZÁLEZ, M. A. S. – *Metodologia de Avaliação de Imóveis – Novo Hamburgo*, SGE, 2003.
- GONZÁLEZ, M. A. S. – *Aplicação de técnicas de descobrimento de conhecimento em bases de dados e de inteligência artificial em avaliações de imóveis – Novo Hamburgo*, SGE, 2003.
- GONZÁLEZ, M. A., Formoso, C. T., “Análise Conceitual das dificuldades na Determinação de Modelos de Formação de Preços através de Análise de Regressão”, *Universidade Federal do Rio Grande do Sul – Núcleo Orientado para a Inovação da Edificação (NORIE)*, Págs. 65 – 75, número 8 – 2000;

- GUEDES, J. C., “Aplicação de Redes Neurais na Avaliação de Bens - Uma comparação com à análise de regressão”, II Simpósio Brasileiro de Engenharia de Avaliações (AVALIAR), São Paulo, Anais do Simpósio, p. 3-10, 2000.
- GUEDES, J. C., “O EMPREGO DE INTELIGÊNCIA ARTIFICIAL NA AVALIAÇÃO DE BENS”, VIII Congresso de Avaliações e Perícias (COBREAP), Florianópolis, 1995.
- GUJARATI, D. N., *Econometria Básica – 3ª Ed.* – São Paulo: MAKRON, 2000.
- HAGAN, M. T., MENHAJ, M. B., “Training Feedforward Networks with the Marquardt Algorithm”, *IEEE Transactions on Neural Networks*, vol. 5, no. 6, 1994.
- HASSIBI, B., Stork, D. G., In *Advances in neural information processing systems 5*, Ed.: Morgan Kaufmann, San Mateo, CA, p 164, 1993.
- HAYKIN, S., “Redes neurais: princípios e prática”, 2º Ed., Porto Alegre: BOOKMAN, 2001.
- HAYKIN, S., “Neural Networks, A Comprehensive Foundation”, Prentice Hall, Upper Saddle River, New Jersey, p 218, 1999.
- IBAPE, Diversos Autores, *Anais do I Congresso Brasileiro de Avaliações e Perícias de Engenharia – São Paulo: PINI, 1978.*
- JOURNEL, A.G., HUIJBREGTS, C.J., “Mining Geostatistics”, Academic Press, New York, 600 p, 1992.
- KITANIDIS, P.K. “Introduction to Geostatistics. Applications in hydrogeology.” Cambridge University Press, 249 p, 1997.
- KMENTA, J., “Elementos de econometria – teoria estatística básica”, V. 1, 2ª Ed., São Paulo: ATLAS, 1988.
- KMENTA. J., “Elementos de econometria – Teoria Econométrica Básica”, V.2, 2ª Ed., São Paulo: ATLAS, 1988.
- KOVÁCS, Z. L., “Redes Neurais Artificiais: Fundamentos e Aplicações”, 3ª Ed., São Paulo: LIVRARIA DA FÍSICA, 2002.
- MATHERON, G., “Lês Variables Regionalisées et leur Estimation, Masson”, Paris, 1965.

- MOREIRA, A. L., “Princípios de Engenharia de Avaliações”, 5a Ed., São Paulo: PINI, 2001.
- NOVAES, L.F.L., “Envoltória sob Dupla Ótica aplicada na Avaliação Imobiliária em Ambiente do Sistema de Informação Geográfica”, Tese de Doutorado em Engenharia de Produção – COPPE / UFRJ - 2002.
- NOVAES, L.F.L., Paiva, S.A., “Situações onde a Aplicação da Análise de Envoltória de Dados (DEA) Apresenta Vantagens Sobre o Uso da Regressão Linear Múltipla”, XII COBREAP, Belo Horizonte / MG, 2003.
- PELLI NETO, A., ZÁRATE, L. E, “Avaliação de Imóveis Urbanos com Utilização de Redes Neurais Artificiais”, XII Congresso Brasileiro de Engenharia de Avaliações e Perícias, COBREAP, Belo Horizonte, Anais COBREAP, Set. 2003^a – Resumo p.14, 2003.
- PELLI NETO, A., “Avaliação de Imóveis Urbanos com utilização de Sistemas Nebulosos (Redes Neuro-Fuzzy) e Redes Neurais Artificiais”, Congreso Panamericano de Valuación - SCdA UPAV, Cartagena / Colombia - 20 a 23 de setembro de 2004, trabalho 08, Versão digital em CD, 2004.
- PELLI NETO, A., ZÁRATE, L.E, “Valuation of Urban Real Estate Through Artificial Neural Networks”, Artificial Intelligence and Applications, September 8 -10, Benalmádena / Spain, A Publication of the International Association of Science and Technology for Development, IASTED, V. 403-169, p. 523-528 , 2003.
- PELLI NETO, A., BRAGA, A. P., “Redes Neurais Artificiais: aplicação e comparação dos resultados com regressão linear na avaliação de imóveis urbanos”, V concurso Internacional de Avaluación y Catastro, SOITAVE, Caracas/ Venezuela, 17 a 20 de Abril de 2005, versão digital em CD, 2005.
- PELLI NETO, A., MORAIS, G, R., “RNA sob dupla ótica – Modelando a análise envoltória de dados (EDO-DEA) para aplicação nas avaliações de imóveis urbanos”, XII COBREAP – Congresso Brasileiro de Avaliações e Perícias, Fortaleza / CE, 2006.
- PUCCINI, A. L., “Matemática financeira – Objetiva e Aplicada”, 6^a Ed., São Paulo: SARAIVA 2002.

- REED, Russel, "Pruning Algorithms – A Survey", IEEE Transactions on Neural Networks, vol. 4, no. 4, 1993.
- RUSSELL, S. J., NORVING, P., "Inteligência Artificial" - 2ª Ed. – Rio de Janeiro: ELSEVIER, 2004.
- SOLLA, S., CUN, Y. L., DENKER, J., "Optimal Brain Damage – In Advances in Neural Information Processing Systems". NIPS 2, 598-605, San Mateo, Morgan Kaufmann Publishers Inc."
- STEWART, J., "Cálculo", Volume I, 4ª Edição - São Paulo: PIONEIRA , 2003;
- STORK D., HASSIBI B., "Second order derivatives for networking pruning: Optimal Brain Surgeon – In Advances in Neural Information Processing Systems". NIPS 5, 164-171, San Mateo, Morgan Kaufmann Publishers Inc.", 1993.
- WERKEMA, M. C. C., AGUIAR, S., "Análise de regressão: como entender o relacionamento entre as variáveis de um processo", V. 7, Belo Horizonte: Fundação Cristiano Ottoni, Escola de Engenharia da UFMG, 1996.