

Redes Neurais Espaciais aplicadas às avaliações em massa Estudo de caso para a cidade de Belo Horizonte / MG

Pelli Neto, Antônio

Eng. Mecânico e Civil, CREA 6.021/D-DF, Caixa Econômica Federal – Entidade Conveniada
Belo Horizonte/MG – Fone/Fax: 3248-5200
antonio-pelli.neto@caixa.gov.br

Resumo. *A construção de modelos de formação de preços imobiliários através 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 RNA's 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 RNA's, às atualmente utilizadas, de forma a aprimorar o processo de avaliação de bens.*

Palavras Chaves: *Engenharia de avaliações, Redes neurais artificiais, Regressão Espacial.*

Abstract. *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 the ANN / RNA's are the best computational tools 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 the ANN / RNA's, rather than those currently in use, so as to improve the real estate evaluation process.*

Key Words: *Engineering of Valuations, Artificial Neural Networks, Spatial Regression.*

1. INTRODUÇÃO

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 do imóvel. 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,

$$P_{(x)} = p(x_1, x_2, \dots, x_n) \quad (1.1)$$

em que P(x) é o preço do bem e p(x₁ a x_n) são os preços de suas características ou atributos relacionados a questões estruturais (físicas), de localização e aspectos econômicos. 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 válida. 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 dos modelos de regressão linear, encorajando os estudiosos a buscar fundamentos em outras metodologias, como as Redes Neurais Artificiais (RNA's).

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

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

² Primeira 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 inicialmente 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 DEA foi inicialmente introduzido no Brasil no Avaliar – Simpósio Brasileiro.

A utilização das RNA'S 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 RNA'S e a Regressão Linear, indicando vantagens desta última abordagem em relação às RNA'S, 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 RNA's, a exemplo de Tay e Ho (1994), que investigaram a aplicabilidade das Redes Neurais Artificiais para as avaliações em massa, para a cidade de Singapura, utilizando o algoritmo *backpropagation*. No Brasil, estudos recentes têm demonstrado boa aceitação das RNA's, a exemplo de Guedes (1995), que propões um estudo comparativo com a Regressão Linear, indicando um melhor desempenho das RNA's e Pelli e Braga (2005) que demonstraram o poder computacional das RNA's, dispensando artifícios matemáticos como as variáveis de interação.

Nos trabalhos desenvolvidos tendo como modelagem as RNA's 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 sofisticados (softwares⁶ 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) introduz e incorpora 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 encontra 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 RNA's e do EDO/DEA.

1.1 OBJETIVOS DO TRABALHO

Este trabalho pretende atingir um objetivo principal, de caráter metodológico, que é demonstrar a possibilidade da utilização das Redes Neurais Artificiais 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).

⁶ No mercado nacional existe o software SisPlanV (2005), que trata a regressão espacial, bem como o software SpaceStat, desenvolvido por Anselin (1998).

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 hedônicos, 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.

2. METODOLOGIAS PARA AVALIAÇÃO DE IMÓVEIS

2.1 INTRODUÇÃO

Nesta seção 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 Redes Neurais Artificiais.

2.2 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 (2.1).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, i = 1, \dots, m \quad (2.1)$$

Onde $Y_i \dots Y_m$ é a variável dependente, explicada ou variável de saída, X_{i1}, \dots, X_{ik} são as variáveis independentes, explicativas ou de entrada, normalmente associadas às características físicas, de localização, e aos aspectos econômicos, β_0, \dots, β_k são denominados de parâmetros da população e $\varepsilon_1, \dots, \varepsilon_m$ são os erros aleatórios do modelo.

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 Mínimos Quadrados Ordinários (MQO). O vetor de coeficientes do modelo, também chamados de vetor de preços implícitos ou hedônicos, é obtido por (2.2).

$$b = (X'X)^{-1} X'Y \quad (2.2)$$

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 Mínimos Quadrados Ordinários, sejam não-tendenciosos, eficientes e consistentes, 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 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 envolvidas sejam adequadamente escolhidas, com o objetivo de garantir a linearidade do modelo. Este modelo é denominado de Modelo Clássico de Regressão (MCR).

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.

- **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 grau de colinearidade. Em Gonzalez (2003), se relacionam as causas das rupturas destes pressupostos básicos da regressão 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 influencia 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 mínimos quadrados são, de fato, inconsistentes. No caso da regressão linear, as soluções a serem obtidas com o método dos mínimos 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 atinge parâmetros satisfatórios de análise; Neste caso, modelos não lineares podem apresentar melhores resultados. Como artifício matemático, utiliza-se das 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, nos modelos lineares;
- Violação de certos pressupostos básicos da regressão, principalmente no que se refere à normalidade, homocedasticidade e auto-regressão, que pode levar a resultados indesejados;
- 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.

2.3 REGRESSÃO ESPACIAL

- INTRODUÇÃO

O valor de mercado de imóveis urbanos sendo estimado pelo Modelo Clássico de Regressão (MCR), 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 MCR 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 recebeu grande impulso nas aplicações realizadas por Anselin (1988), com o desenvolvimento da ferramenta computacional denominada SpaceStat (Anselin, 1990).

No Brasil, o primeiro trabalho 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.

Ocorre que, quando se trata de dados distribuídos espacialmente, como é o caso de dados habitacionais, podem existir erros de medidas em relação à exata localização do imóvel, como também efeitos de interação espaciais. Estas razões são suficientes para que o pesquisador considere um fator adicional no 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 (2.1) são ineficientes. Neste caso, testes de significância como t de Student e F de Snedecor, bem como os intervalos de confiança inferidos, não são mais válidos e as decisões tomadas com base neles são enganosas (Dantas 2001). Da mesma forma, 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 (2.1). Em ambos os casos, o MCR mostra-se inadequado, devendo ser substituído pelos Modelos Espaciais, estimados por uma nova metodologia denominada Modelagem por Econometria Espacial.

- MODELAGEM POR ECONOMETRIA ESPACIAL

A Regressão Espacial trata do efeito causado pela 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 (2.1) e que será apresentado na seção 2.3.3, 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 (2.1), que será mostrado na seção 2.3.4.

Existem duas formas 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 LM⁷ Robusto (erro) e LM Robusto (defasagem). No primeiro caso, a inferência espacial é realizada pelo processo denominado de Krigeagem, desenvolvido por Matheron (1965); no segundo caso, a modelagem espacial é realizada conforme a metodologia desenvolvida por Anselin (1988), que é apresentada a seguir.

- **MODELAGEM PELA METODOLOGIA DESENVOLVIDA POR ANSELIN**

Para diagnosticar a presença de efeitos de dependência espacial, bem como introduzir estes efeitos no modelo (2.1), 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. Em geral, a matriz de W é padronizada por linha, assumindo a nomenclatura W_s . Neste caso, cada elemento de W_s , representado por W_{sij} , é 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 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.

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 forma resumida. Maiores detalhes podem ser encontrados Anselin (1988a). É importante frisar que a validade destes testes exige a aceitação das hipóteses de normalidade e homocedasticidade dos resíduos de MQ, obtidos pelo modelo (2.1).

Uma vez detectada a presença de autocorrelação espacial nos dados, faz-se necessário introduzir extensões convenientes no modelo tradicional, representado na equação (2.1), considerando-se os efeitos autocorrelação espacial nos erros, através do Modelo de Erro Espacial, e os efeitos ocasionados pelas interações entre os preços, pelo Modelo de Defasagem Espacial, como mostra-se a seguir.

⁷ LM é a sigla de Multiplicador de Lagrange.

- 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 de uma cidade, que não necessariamente coincidem com a verdadeira dimensão do fenômeno observado. Isto é, na prática, o consumidor não tem o conhecimento exato dos limites que dividem os bairros. No mercado imobiliário 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. Outro fator que pode gerar a autocorrelação espacial nos erros é a omissão de variáveis locais relevantes.

Para tratar adequadamente este tipo de efeito espacial nos dados, a primeira modificação com relação à equação (2.1) será considerar o processo espacial autoregressivo no termo de erro, da seguinte forma:

$$\varepsilon = \lambda \mathbf{W}\varepsilon + u \quad \text{ou} \quad \varepsilon = (\mathbf{I} - \lambda \mathbf{W})^{-1} u, \quad (2.3)$$

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 (2.3) em (2.1) resulta no seguinte modelo de erro espacial:

$$Y = X\beta + (\mathbf{I} - \lambda \mathbf{W})^{-1} u \quad (2.4)$$

Para estimações eficientes dos parâmetros do modelo (2.4) é necessário usar o estimador de verossimilhança, que consiste em maximizar a função de log-verossimilhança dada por (2.5), utilizando-se técnicas de otimização não linear.

$$L = \frac{n}{2} \ln(\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \lambda \mathbf{W}| - \frac{1}{2\sigma^2} \varepsilon' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) \varepsilon, \quad (2.5)$$

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 (2.3).

Como comentado na seção 2.3.3, quando os erros são autocorrelacionados espacialmente, os parâmetros estimados pelo Modelo Tradicional (2.1) são não eficientes, isto é, os desvios padrões que se encontram associados a eles são tendenciosos. Assim, os testes t e F , como também os intervalos de confiança construídos não são mais válidos e os resultados obtidos a partir deles são enganosos.

- O MODELO DE DEFASAGEM ESPACIAL

O efeito de defasagem espacial é ocasionado pela dependência espacial criada através de uma interação espacial entre os preços dos imóveis, uma espécie de “efeito de vizinhança” na determinação dos preços. Isto é, 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 (2.1), dada por WY, que é a variável dependente espacialmente defasada. Cada elemento w_{yi} , 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, educação, etc. A introdução do termo de defasagem espacial, como variável independente, serve como “proxy” para as variáveis independentes omitidas que estão correlacionadas com as características locacionais (Pace, Barry e Sirmams, 1998). Com a incorporação desta variável, o modelo (2.1) passa a ser

$$Y = X\beta + \rho WY + \varepsilon, \quad (2.6)$$

onde ρ é o coeficiente de autocorrelação espacial da variável WY, ε é idêntica e independentemente distribuído.

Tendo em vista que a variável WY é aleatória, a estimação por mínimos quadrados ordinários não é adequada, porque viola um dos pressupostos básicos do Modelo Clássico de Regressão⁸. Observe-se também que, ao comparar os modelos (2.1) com (2.6), constata-se no primeiro a falta da variável WY, o que gera um grave erro de especificação⁹. Neste caso, as avaliações realizadas por (2.1) são tendenciosas e inconsistentes. Da mesma forma que no modelo (2.4), a estimação deve ser realizada pelo método da máxima verossimilhança, que consiste na maximização da função (2.7) utilizando técnicas de otimização não linear.

$$L = \frac{n}{2} \ln(\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |\mathbf{I} - \rho \mathbf{W}| - \frac{1}{2\sigma^2} \varepsilon' \varepsilon \quad (2.7)$$

- 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 de LM – Robutso e LM – Defasagem. 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).

2.4 REDES NEURAIIS ARTIFICIAIS

Nesta seção procurou-se dar mais ênfase à parte conceitual do que formal, com o objetivo de permitir uma utilização do texto como uma primeira leitura para os profissionais de Engenharia de Avaliações que estejam ingressando nesta área. Apesar de ter sido minimizada, em algumas seções se fará uso da parte formal, para evitar prejuízos aos interessados no estudo das RNA's.

Os estudos de Redes Neurais Artificiais são relativamente novos. De 1943, quando da publicação do artigo de autoria de *Warrem McCulloch*¹⁰ e *Walter Pitts*¹¹, até hoje, muitas pesquisas vem sendo realizadas em todos os âmbitos da ciência, com a medicina,

⁸ As variáveis explicativas não devem conter nenhuma perturbação aleatória

⁹ No modelo devem estar presentes todas as variáveis explicativas relevantes.

¹⁰ Warrem McCulloch - Psiquiatra e Neuroanatomista

¹¹ Walter Pitts - Matemático

biologia e engenharia. Não existem livros técnicos que tratem com especificidade as Redes Neurais Artificiais – RNA’S e sua aplicação na Engenharia de Avaliações. Porém, podemos 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, nos forneceram subsídios para a aplicação desta metodologia e hoje temos como resultado a citação das Redes Neurais como metodologia científica reconhecida na NBR 14653, Avaliação de Bens, Parte 2 – Imóveis Urbanos.

- BREVE HISTÓRICO DAS REDES NEURAIIS 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 *Warrem McCulloch*¹² e *Walter Pitts*¹³. 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.

Em 1949, Donald Hebb¹⁴ 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 RNA’s através de circuitos analógicos e, naquela época, acreditou-se que o caminho para o entendimento da inteligência humana havia sido descoberto. Nathaniel Rochester¹⁵ (1956) desenvolveu uma simulação em computador do neurônio de McCulloch & Pitts, com regra de treinamento Hebbiana.

Frank Rosenblatt¹⁶ (1957) desenvolveu o Perceptron, que tinha como objetivo o reconhecimento de padrões ópticos (modelo da visão humana). Em 1958 Rosenblatt introduziu o primeiro modelo de rede neural artificial, estabelecendo a base para a Inteligência Artificial.

Bernard Widrow¹⁷ desenvolveu um novo tipo de elemento de processamento de redes neurais chamado de Adaline, equipado com uma poderosa lei de aprendizado, que diferente do Perceptron, ainda possui aplicabilidade na atualidade. Fundou a primeira empresa de circuitos neurais digitais, a Memistor Corporation.

Marvin Minsky¹⁸ desenvolveu um estudo contrário à lógica. Escreveu o livro Perceptron, onde demonstrava as limitações da Inteligência Artificial. Em uma rigorosa análise matemática ficou 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 RNA’s, como o proposto por John Hopfield¹⁹ e Teuvo

¹² Warrem McCulloch - Psiquiatra e Neuroanatomista

¹³ Walter Pitts - Matemático

¹⁴ Donald Hebb – Biólogo – Estudava o comportamento dos animais

¹⁵ Nathaniel Rochester – IBM produziu um dos primeiros programas de IA

¹⁶ Frank Rosenblatt – Pesquisador Norte Americano – (1928-1969)

¹⁷ Bernard Widrow – Cientista – Criador do Adaline

¹⁸ Marvin Minsky –Americano - Professor da Universidade de Carnegie-Mellon. Um dos pioneiros nos estudos da Robótica.

¹⁹ John Hopfield – físico e biólogo professor da Universidade de Princeton.

*Kohonen*²⁰. Finalmente em 1986, *David Rumelhart*²¹ 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 RNA's. Os engenheiros da computação forneceram os artefatos que tornaram possíveis as aplicações da Inteligência artificial.

- **REDES NEURAIAS 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 colegas pesquisadores já afirmavam a importância deste novo conceito, desenvolvendo pesquisas nesta área, sendo, aliás, responsáveis pelos avanços que culminaram com a aceitação das RNA's 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 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 Engenharia²².

- **CONCEITOS BÁSICOS**

As RNA's, ou redes neurais artificiais, foram desenvolvidas a partir de uma tentativa de reproduzir em computador um modelo computacional que simule a estrutura e funcionamento do cérebro humano. Uma RNA é uma implementação de um algoritmo que tem como base o funcionamento do cérebro humano. As RNA's 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 entre os elementos de processamento e se a sua natureza é excitatória ou inibitória. A rede de neurônios está disposta em camadas responsáveis pelas entradas das informações (camada de entrada – correspondendo às variáveis independentes utilizadas no mercado imobiliário), processamento destas informações (camada intermediária) e a produção de resultados (camada de saída – que corresponde às variáveis dependentes, normalmente valor unitário ou valor total).

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.

²⁰ Teuvo Kohonen – Professor acadêmico da Finlândia, especializado em Memórias Associativas

²¹ David Rumelhart – Estudioso da psicologia cognitiva

²² www.pellisistemas.com.br

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. De forma conceitual, podemos considerar que as RNA's são modelos matemáticos que se assemelham à estrutura do cérebro humano e possuem capacidade de aprendizagem para posterior generalização.

- **CLASSIFICAÇÃO DOS MODELOS DE REDES NEURAIS ARTIFICIAIS**

Os modelos de RNA's podem ser classificados em diversos grupos, de acordo com suas características e propriedades. Estes grupos dividem as redes neurais de acordo com o tipo de treinamento – que pode ser auto-aprendizado ou aprendizado supervisionado. O treinamento supervisionado consiste em apresentar à rede neural um padrão a ser reconhecido juntamente com a resposta que a rede deve fornecer ao deparar-se novamente com o mesmo padrão. Na engenharia de avaliação, os padrões a serem apresentados na entrada são os elementos amostrados, constituídos pelas variáveis de entrada e a saída da rede é o valor unitário ou valor total. No treinamento com auto-aprendizado os padrões são apresentados para a rede e esta se encarrega de agrupar aqueles que possuem características similares. Este processo é também chamado de clusterização²³. As regras de aprendizado são formuladas através do algoritmo competitivo ou algoritmo adaptativo por correção de erros. As redes neurais que utilizam regras de aprendizado do tipo competitivas caracterizam-se pelas conexões laterais dos neurônios com seus vizinhos, estabelecendo, assim, uma competição entre os neurônios. As redes de Hopfield e Kohonen se enquadram nesta categoria. As redes com aprendizado do tipo correção de erros são baseadas no princípio da adaptação e correção dos pesos de atuação de cada neurônio, até que este produza a saída desejada. A correção de erros está diretamente ligada ao aprendizado do tipo supervisionado. Este é o caso mais comum para a aplicação na engenharia de avaliação, para as avaliações comparativas do mercado imobiliário.

De forma geral, as etapas do treinamento são compreendidas pela inicialização dos pesos com valores dentro de determinado intervalo (normalmente inicializados aleatoriamente no intervalo compreendido entre +- 0,5). Após o processo de inicialização, é apresentado o vetor de entrada (variáveis independentes que representam os atributos dos elementos amostrados) com as respectivas saídas desejadas (valor unitário ou total). Este processo continua com a apresentação de novas entradas até que todos os padrões sejam apresentados e o erro esteja tão pequeno quanto se deseja.

Assim como é difícil para muitos especialistas expressar formalmente o seu conhecimento de forma a ser usado em um programa de computador, extrair o conhecimento das Redes Neurais Artificiais (que são modelos baseados no funcionamento do cérebro humano) representa também um desafio. Tal problema assemelha-se ao de transformar um grande conjunto de dados (imagens, sons, sinais de sensores elétricos e mecânicos, etc.) em fórmulas que abstraíam tal conhecimento de uma maneira simples e inteligível para o usuário. Tal ferramenta torna-se então indispensável em qualquer sistema de Prospecção de Dados (“Data Mining”) e de descoberta automática de conhecimento (“Knowledge Data Discovery”).

As RNA's são capazes de resolver, basicamente, problemas de aproximação, ou regressão. As funções de ativação usualmente escolhidas são a tangente hiperbólico para os neurônios da camada escondida e a linear para a saída da rede.

²³ Clusterização – Divisão de Classes

O objetivo principal do aprendizado em Redes Neurais é a obtenção de modelos com relativa capacidade de generalização, tendo como base o conjunto de dados.

Duas características importantes das RNA'S são o conhecimento que é adquirido pela rede através de um processo automático de aprendizado e a intensidade da ligação entre os neurônios, conhecidos como pesos sinápticos, que são utilizados para armazenar o conhecimento.

Tais características fazem das RNA's uma ferramenta fundamental para automatizar o processo de aquisição do conhecimento. Entretanto, devido à natureza distribuída das Redes Neurais, expressar este conhecimento de uma forma simples torna-se uma tarefa difícil.

- **REDES NEURAIAS 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 (*pruning*) 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 para a poda de redes neurais: 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 rede neural não é retreinada após os cortes. No método OBS, as conexões são cortadas e, após o corte de uma conexão, a rede é retreinada, permitindo que um número maior de cortes seja efetuado. Além disso, no método OBS a rede neural 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 nos permite obter modelos com pequena possibilidade de sobreajuste (*overfitting*). Este fato pode ser observado comparando os resultados obtidos na determinação dos avaliandos utilizando-se redes neurais sem poda e com poda conforme será mostrado no próximo capítulo.

2.5 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, estudamos 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 Redes Neurais Artificiais, 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 rede ser acrescida a variável de defasagem espacial, a ser abordada no próximo capítulo.

3. ESTUDO DE CASO – CIDADE DE BELO HORIZONTE

3.1 INTRODUÇÃO

Na seção anterior foram apresentados os tratamentos científicos, a serem utilizados neste trabalho, descritos na NBR 14.653-2, e que conseqüentemente podem ser empregados para as modelagens de dados de mercado. Nesta seção 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 Redes Neurais Artificiais.

Para o diagnóstico de dependência espacial serão utilizados variogramas e testes de autocorrelação espacial, conforme metodologia definida por Anselin (1998). Para o tratamento dos dados serão utilizados os programas SisPlanV²⁴, para a Regressão Espacial e a análise de autocorrelação espacial, o SisReN²⁵ para a modelagem das Redes Neurais Artificiais e para o processo de poda das estruturas das redes neurais artificiais e EASYKRIG²⁶ versão 3.0 (toolbox para Matlab) para a construção dos variogramas.

3.2 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 (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, quarteirões simétricos, 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 erradas. 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. Ficou comprovado que o terreno da cidade era mais seco, portanto não necessitava de prévia drenagem. As condições 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 capital do Estado. O prazo mínimo para a transferência definitiva do governo era de 4 anos. O tempo foi insuficiente e a cidade teve que ser inaugurada às pressas, ainda poeirenta e com prédios a construir. Sua consolidação levou anos.

²⁴ Sistema de Regressão Linear e Regressão Espacial desenvolvido pela Pelli Sistemas Engenharia Ltda.

²⁵ Sistema de Regressão Linear e de Redes Neurais Artificiais desenvolvido pela Pelli Sistemas Engenharia Ltda.

²⁶ Toolbox para o Matlab para modelagem de dados com a Krig, permitindo a construção de variogramas.

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, conforme TABELA 3.1:

TABELA 3.1 - 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
BR	169.799.170	8.514.215,30	19,94

Fonte: IBGE. Censo Demográfico 2000

Na TABELA 3.2 está a quantidade de domicílios catalogados no censo de 2000:

TABELA 3.2 - 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

A administração pública municipal está dividida nos setores da FIGURA 3.1:



FIGURA 3.1 – Regiões administrativas 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 as Prefeituras Municipais dos grandes municípios brasileiros, para a construção e manutenção de modelos científicos que descrevam com eficiência o mercado imobiliário local.

3.3 PESQUISA DE DADOS

A pesquisa de dados no município de Belo Horizonte teve início em junho 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 SisDat – 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 construir a matriz de pesos, indicando as distâncias calculadas no plano cartesiano, em metros lineares, entre cada dado do banco de dados. Esta matriz de pesos foi utilizada na Regressão Espacial e nas Redes Neurais Artificiais.

3.4 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 a serem aplicados aos dados, optamos por utilizá-las, sem alteração de suas escalas. A TABELA 3.3 contém uma sumarização da amostra com a especificação das variáveis utilizadas na modelagem, apresentando os valores mínimos, médios, máximos e a amplitude.

TABELA 3.3 – Descrição da amostra utilizada para estimação dos modelos de preços hedônicos

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

Os resultados apresentados na TABELA 3.3 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, 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 até aqueles com idade de 40 anos, com conservação variando de ruim a boa e padrão de mínimo a luxo.

3.5 TRATAMENTOS DOS DADOS

- 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 3.1,

$$P = f(L, E, T, \beta) + \varepsilon \quad (3.1)$$

onde f é indicativo da forma funcional, β são os parâmetros e ε são os resíduos ou erros aleatórios. Como os dados são extremamente heterogêneos, e buscando evitar a construção de modelos heterocedásticos, foram eliminados da modelagem todos os dados com preço unitário acima de R\$ 2.000,00 R\$/m².

Para a estimação empírica do modelo 3.1 será utilizado como uma primeira abordagem o Modelo Clássico de Regressão Linear, via Mínimos Quadrados Ordinários (MQ). Nesta abordagem não está sendo 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 3.3. Pelos resultados obtidos na TABELA 3.4, verifica-se que as hipóteses básicas da normalidade, homocedasticidade dos erros aleatórios foram inicialmente aceitos, considerando nesta etapa o modelo como consistente para a avaliação de apartamentos no mercado de Belo Horizonte. Durante o processo de modelagem, foram aplicadas as transformações ln (logaritmo neperiano) e a função inversa, em todas as variáveis, buscando um modelo que apresente o maior poder de predição, e dentre os modelos disponibilizados foi selecionado o que apresentou o melhor coeficiente de correlação.

TABELA 3.4 – Resultados da estimação dos modelos 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 modelo inferido permitir 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 0,57% ao mês. 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 F de calculado foi 53,71, com a hipótese nula rejeitada a menos do que 0,01%. O coeficiente de correlação foi de 0,8568, indicando uma forte correlação entre as variáveis independentes e a variável dependente e a distribuição dos resíduos sinaliza que o modelo possui tendências à normalidade, conforme observa-se no GRÁFICO 3.1,

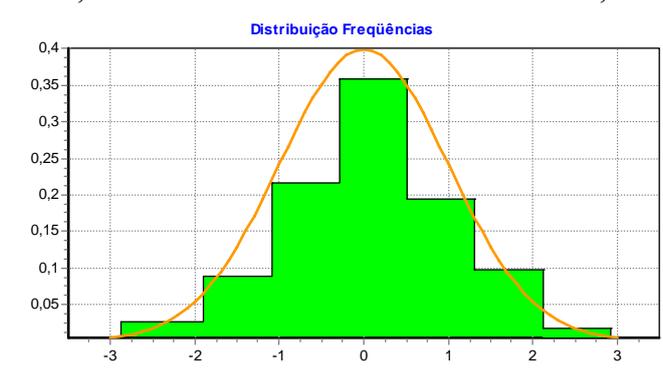


GRAFICO 3.1 – Distribuição de Frequência dos resíduos do MCR

O modelo apresentou um bom poder de predição, contudo apresenta uma tendência de subavaliar imóveis de valores mais altos, que pode ser verificado no GRÁFICO 3.2, que apresenta os preços por valores estimados do modelo,

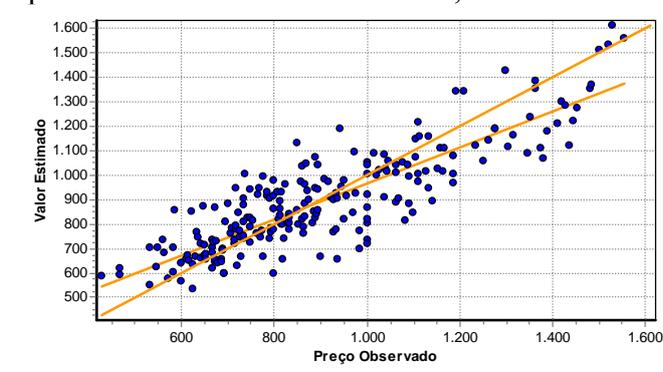


GRAFICO 3.2 – Valores estimados por Preços observados

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

$$\begin{aligned} \text{Valor Unitário} = & e^{(+7,039344981+0,0008226328371 * \text{Setor Urbano}+0,1613379329 * \\ & \ln(\text{Nivel/Elevador})+0,005733353163 * \text{Data do Evento}+0,001348528869 * \\ & \text{Total Vagas}-0,5026981916 * \ln(\text{Área Coberta})+0,002915663288 * \text{Área Desc.}+ \\ & 0,1168737474 * \text{Dormitórios} + 0,009908174092 * \text{Equipamentos}- \\ & 0,05386150912 * \ln(\text{Unidades por Acesso})+0,001020533186 * \text{Padrão} + 0,0009130749008 \\ & * \text{Conservação)} \end{aligned} \quad (3.2)$$

Os resultados obtidos com o MCR via Mínimos Quadrados Ordinários apresenta resultados estatísticos razoáveis, 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.

• REGRESSÃO ESPACIAL

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 acrescida a dimensão espacial. A rigor, antes de proceder aos testes de inferência espacial, não se pode concluir que o modelo MCR, 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 3.3,

$$P = f(L, E, T, \beta, WP) + \varepsilon \quad (3.3)$$

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 MCR da seção anterior.

Para a estimação empírica do modelo 3.3 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 variograma e determinação da distância máxima de influência do preço de um imóvel em relação a seus vizinhos. O Gráfico 3.3 abaixo indica a posição dos dados no espaço urbano, considerando a latitude e longitude em coordenadas UTM, mostrando a região do Bairro Belvedere como a de maior valorização imobiliária, cujos preços estão na ordem de 3.000 R\$/m².

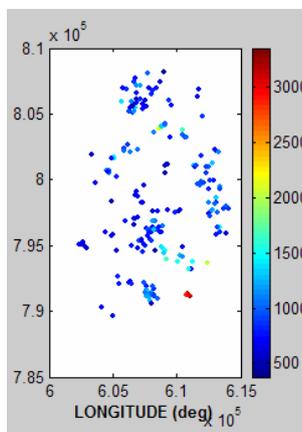


GRAFICO 3.3 – Dados de mercado (latitude x longitude)

Em seguida, foi construído o variograma, cujos resultados podem ser obtidos no GRÁFICO 3.4 a seguir, onde ficou constatado a influencia de um evento de mercado em relação a sua vizinhança em um raio 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.

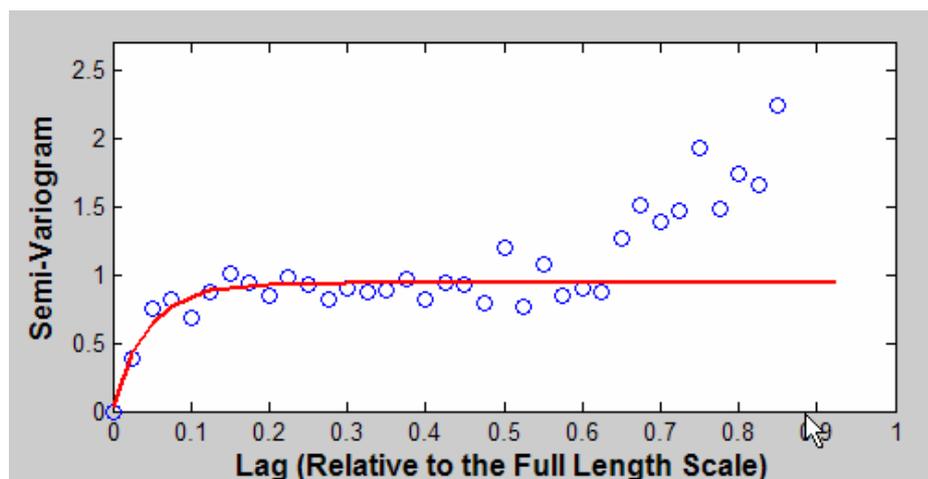


GRAFICO 3.4 – Variograma para os dados de mercado de Belo Horizonte / MG

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 3.4. 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) indica um forte efeito de defasagem espacial nos preços de habitação, uma vez que se mostrou significativo a menos de 0,01%, enquanto que pelo teste LM Robusto (erro) a hipótese de autocorrelação nos resíduos é rejeitada a um nível de significância de 1,65%. Ou seja, o efeito vizinhança parece estar afetando os preços de habitação na cidade de Belo Horizonte / MG.

Tabela 3.5 Diagnóstico da Dependência Espacial de preços por habitação

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 3.3 utilizando o modelo funcional da equação 3.3. Pelos resultados obtidos na TABELA 3.6, verifica-se que as hipóteses básicas da normalidade, homocedasticidade dos erros aleatórios foram inicialmente aceitos, considerando o modelo como consistente para a avaliação de apartamentos no mercado de Belo Horizonte.

TABELA 3.6 – Resultados da estimação dos modelos de preços hedônicos com a regressão linear

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
WY	X	0,262942164	3,696	0,02
Preços Unitários	Ln(Y)	+7,039344981		

Assim, pode-se concluir que os preços dos imóveis não dependem somente das variáveis explicativas utilizadas pelo MCR, 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 mensal dos imóveis em 0,55% ao mês. 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 F calculado foi de 56,25, superior ao MCR, e a hipótese nula rejeitada a um nível de significância inferior a 0,01%. O coeficiente de correlação foi de 0,8646, superior ao do MCR, o que indica uma forte correlação entre as variáveis independentes e a variável dependente e a distribuição dos resíduos sinaliza que o modelo possui tendências à normalidade, conforme observa-se no GRÁFICO 3.5,

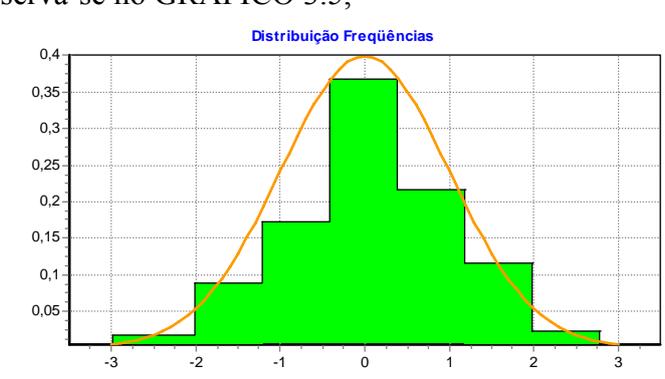


GRAFICO 3.5 – Distribuição de Frequência dos resíduos do MCR

O modelo apresentou um bom poder de predição, melhor do que dos MCR, que pode ser verificado no GRÁFICO 3.6, que apresenta os preços por valores estimados do modelo,

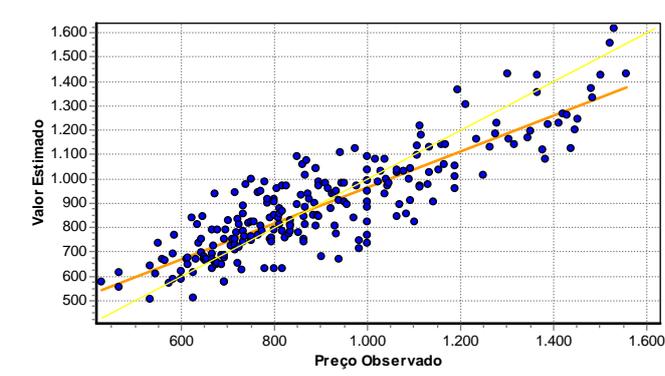


GRAFICO 3.6 – Valores estimados por Preços observados

Os resultados obtidos com a Regressão Espacial foram satisfatórios, tendo sido considerada a existência de autocorrelação espacial. Os resultados da estimação espacial apresentados na tabela 3.6 indicam que o coeficiente ρ , da variável defasada espacialmente $W \ln(P)$ é positivo e estatisticamente significativo 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 3.4 a seguir:

$$\text{Valor Unitário} = e^{(+5,343664799 + 0,0007066240486 * \text{Setor Urbano} +$$

$$\begin{aligned}
 & 0,1787713532 * \ln(\text{Nível/Elevador}) + 0,00551386571 * \text{Data} + \\
 & 0,00129687329 * \text{Total Vagas} - 0,5061941919 * \ln(\text{Área Cob}) + \\
 & 0,002978794592 * \text{Área Desc.} + 0,1208642976 * \text{Dormitórios} + \\
 & 0,009361433193 * \text{Equipamentos} - 0,05063036011 * \ln(\text{Unidades por Acesso}) + \\
 & 0,0009564062885 * \text{Padrão Acabamento} + \\
 & 0,0008277902485 * \text{Estado Conservação} + 0,2629421637 * w * y) \quad (3.4)
 \end{aligned}$$

- 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 Redes Neurais Artificiais. Neste momento não será acrescida a dimensão espacial. A forma do modelo fica definida como na equação 3.5,

$$P = f(L, E, T, w) + \varepsilon \quad (3.5)$$

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 MCR da seção 3.5.1.

Para a estimação empírica do modelo 3.4 será utilizada as RNA's, via algoritmo de treinamento *Levenberg Marquadt*. A estrutura da rede será composta por 6²⁷ (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 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 3.3.

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

Os resultados obtidos com a modelagem estão listados na tabela 3.7:

TABELA 3.7 – Resultados da estimação com as RNA's

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	

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 0,50% ao mês. 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

²⁷ Definido inicialmente de forma empírica.

seja, a localização pode representar até 50% do valor do imóvel, mesmo valor encontrado pelo MCR.

Os pesos da camada intermediária e da camada de saída foram calculados pelo processo de treinamento e estão listados na TABELA 3.8, que inclui o BIAS em ambas as camadas.

TABELA 3.8 – Pesos das camadas intermediárias e da camada de saída

W1											
-0,112	0,212	0,402	-0,218	0,319	-0,165	0,323	-0,031	-0,569	1,025	0,747	
-0,210	-0,371	0,611	0,072	-0,322	-0,825	-0,049	0,108	-0,444	0,036	1,032	
-0,499	0,028	0,222	-0,346	-0,071	-0,857	0,274	0,387	0,063	-0,104	0,399	
-0,010	0,113	-0,356	0,089	0,469	0,175	-0,088	-0,132	0,089	0,203	0,198	
-0,027	0,027	-0,167	-1,077	0,342	-1,065	0,233	-0,210	0,592	0,580	0,236	
0,719	-0,159	-0,345	0,134	-0,811	-0,131	-0,289	-0,211	0,571	0,155	0,571	

w2						
0,956	-0,914	0,767	-0,596	-1,071	0,880	-0,023

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 3.7,

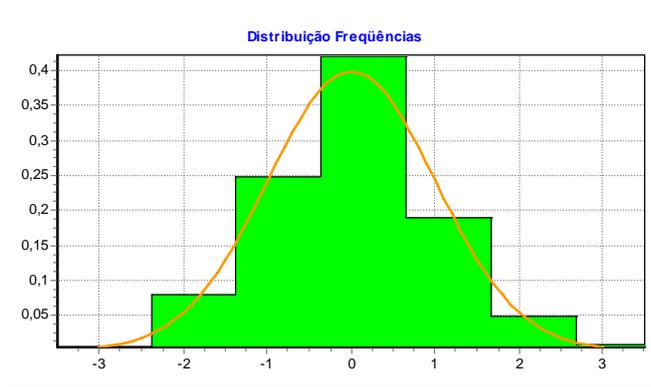


GRAFICO 3.7 – Distribuição de Frequência dos resíduos de RNA's

O modelo apresentou um bom poder de predição, melhor do que da Regressão Espacial e da Regressão Linear, que pode ser verificado no GRÁFICO 3.8, que apresenta os preços observados por valores estimados do modelo.

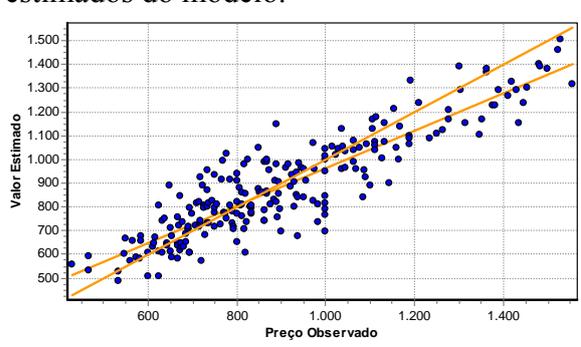


GRAFICO 3.8 – Valores estimados por Preços observados

Os resultados obtidos com as Redes Neurais foram mais satisfatórios do que os demais, apesar de não ter sido considerada a existência de autocorrelação espacial. A aparente superioridade do modelo de RNA's 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.

- REDES NEURAIAS 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 de forma a ser modelado a partir da utilização das Redes Neurais Artificiais. Neste momento será acrescida a dimensão espacial. A forma do modelo fica definida como na equação 3.6,

$$P = f(L, E, T, w, WP) + \varepsilon \quad (3.6)$$

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 MCR da seção 4.5.1, incluindo a variável de defasagem espacial.

Para a estimação empírica do modelo 3.6 será utilizada as RNA's, via algoritmo de treinamento *Levenberg Marquadt*. A estrutura da rede será composta por 5²⁸ (cinco) 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 3.3.

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

Os resultados obtidos com a modelagem estão listados na tabela 3.8:

TABELA 3.8 – Resultados da estimação com as RNA's

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 0,7% ao mês. 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 1.000,00 R\$/m², ou seja, a localização pode representar até aproximadamente 43% do valor do imóvel, valor aproximado ao encontrado pela Regressão Espacial.

²⁸ Definido inicialmente de forma empírica.

Os pesos da camada intermediária e da camada de saída foram calculados conforme a TABELA 3.9, que inclui o BIAS em ambas as camadas.

TABELA 3.9 – Pesos das camadas intermediárias e da camada de saída

W1												
-0,203	-0,360	-0,495	-0,716	-0,313	-0,040	-0,538	-0,235	0,554	-0,721	-0,032	0,449	0,107
-0,257	0,241	-0,251	-1,287	-0,345	-1,168	0,475	0,001	0,458	0,550	0,081	0,283	-0,649
0,272	0,472	0,099	-0,506	-0,296	-0,233	0,456	0,123	0,298	0,673	-0,247	0,394	-0,143
-0,180	-0,025	0,184	-0,171	0,592	-0,303	-0,011	-0,110	0,233	-0,780	-0,538	-0,156	1,067
0,099	-0,303	-0,281	-0,116	0,250	0,340	-0,542	-0,170	0,409	0,080	-0,060	0,202	0,676
-0,203	-0,360	-0,495	-0,716	-0,313	-0,040	-0,538	-0,235	0,554	-0,721	-0,032	0,449	0,107

w2					
0,903	-1,300	0,637	-0,809	-0,809	0,575

O coeficiente de correlação foi de 0,8780 inferior ao das redes neurais artificiais sem a variável de defasagem, mas que indica uma forte correlação entre as variáveis de entrada e a variável de saída. A distribuição dos resíduos sinaliza que o modelo possui tendências à normalidade, conforme se observa no GRÁFICO 3.9,

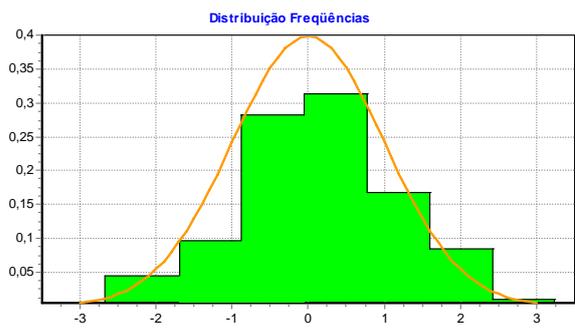


GRAFICO 3.9 – Distribuição de Frequência dos resíduos de RNA's

O modelo apresentou um bom poder de predição, melhor do que da Regressão Espacial e da Regressão Linear, que pode ser verificado no GRÁFICO 3.10 que apresenta os preços observados por valores estimados do modelo.

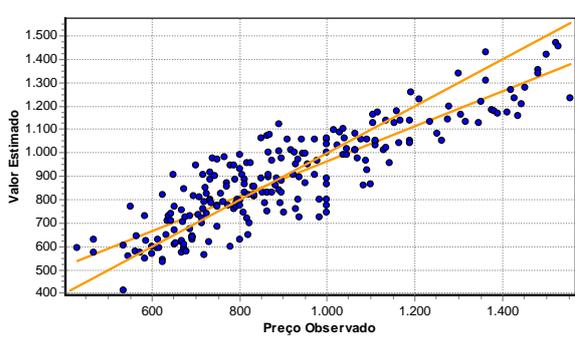


GRAFICO 3.10 – Valores estimados por Preços observados

Os resultados obtidos com as Redes Neurais 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 RNA's 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.

- VALIDAÇÃO DOS RESULTADOS

Considerando que foi elaborada apenas uma modelagem para a comparação entre os diversos tratamentos efetuados, e como não foi introduzido algum processo de validação dos resultados, nesta seção será tratado da validação cruzada.

Para realizar a validação cruzada, com o objetivo de escolher o melhor modelo para as avaliações no mercado imobiliário 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. Na verdade, a amostra inicialmente coletada, composta de 255 dados de mercado, foi subdividida em cinco amostras, cada uma com dois conjuntos de dados, sem um para treinamento e outro para a 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, foram coletados 255 dados no período, sendo 231 dados efetivamente utilizados (quando 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%. Os duzentos e cinquenta e cinco dados georreferenciados foram plotados através do software MapInfo, produzindo o mapa da FIGURA 3.2:

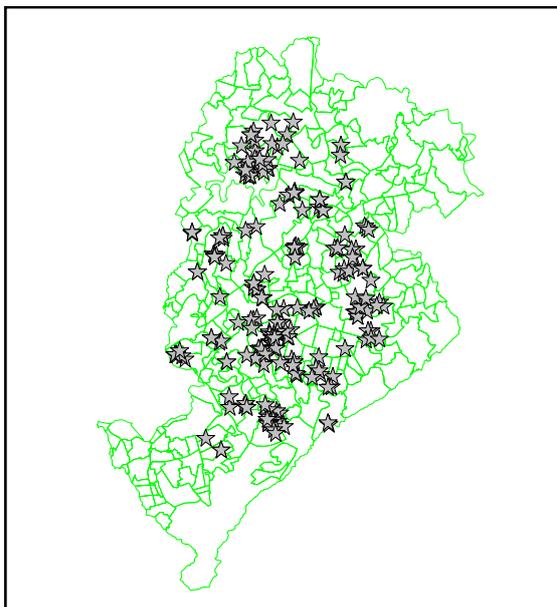


FIGURA 3.2 – Dados coletados no Mercado de Belo Horizonte / MG

A modelagem de dados, utilizando as Redes Neurais Artificiais foi desenvolvida utilizando o software SisReN – Sistema de Regressão Linear e Redes Neurais Artificiais e para as técnicas de poda foi utilizado o Matlab. Para cada uma das 5 amostras foram realizadas cinco modelagens: a primeira com a regressão espacial, utilizando o modelo de defasagem espacial; a segunda, utilizando a regressão linear (MCR); a terceira utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão linear, sem aplicar as técnicas de poda; a quarta utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão linear, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE); e

por ultimo, utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão espacial, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE).

3.6 ANÁLISE DOS RESULTADOS OBTIDOS NAS MODELAGENS

Para a obtenção dos resultados, os mesmos passos foram utilizados, para cada modelagem, conforme descritos na seção 4.5. A diferença esta na aplicação das técnicas de PODA, quando a rede foi retreinada com base na apresentação da amostra de validação, para a remoção de conexões.

Na tabela 3.10 foram listados os resíduos quadráticos médios na validação dos resultados, conforme 3.7,

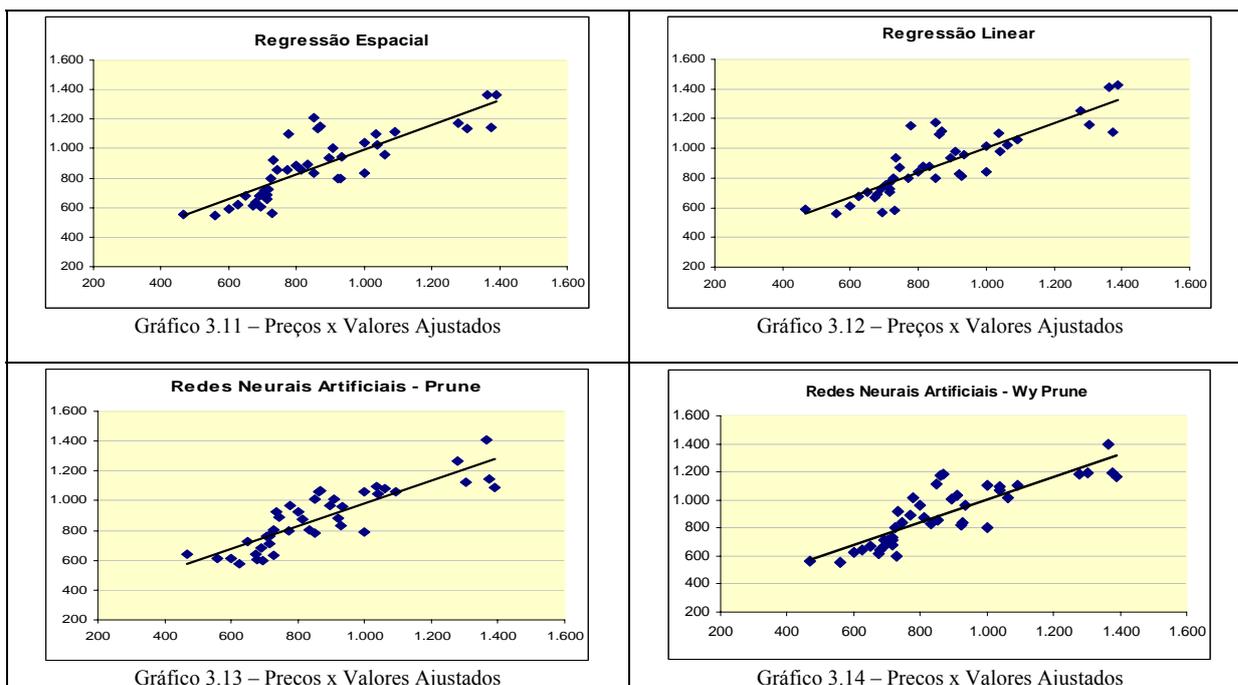
$$\varepsilon = \frac{\sum(P-V)^2}{n} \quad (3.7)$$

Onde ε é igual ao resíduo quadrático médio na validação dos modelos, P os preços dos imóveis que compõe a amostra da validação e V os valores estimados para estes mesmos dados.

TABELA 3.10 – Resultados do processo de validação

	Resíduo Quadrático Médio para a validação				
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

Para melhor visualizar os resultados, os gráficos de 3.11 a 3.58 apresentam os Preços x Valores Ajustados para cada uma das amostras.



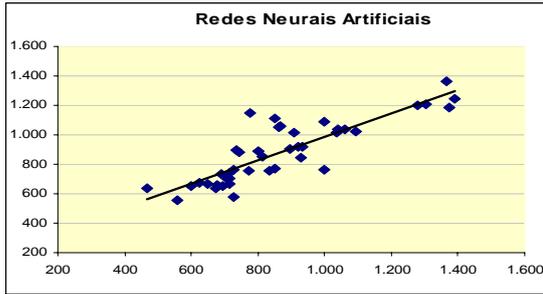


Gráfico 3.15 – Preços x Valores Ajustados

Os gráficos de números 3.11 a 3.15 representam os resultados para a amostra 1.

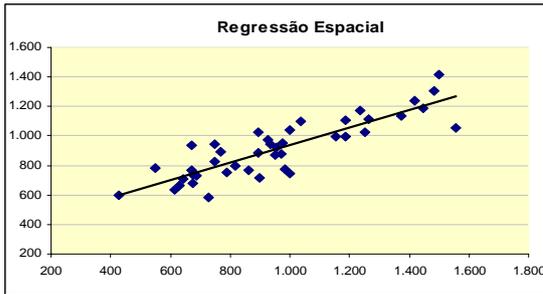


Gráfico 3.16 – Preços x Valores Ajustados

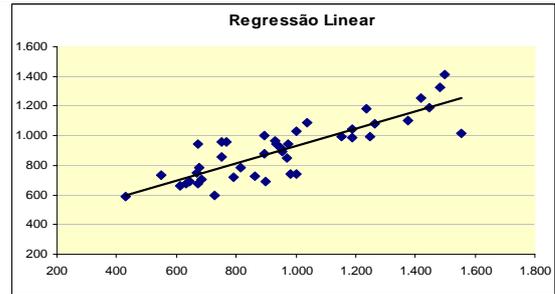


Gráfico 3.17 – Preços x Valores Ajustados

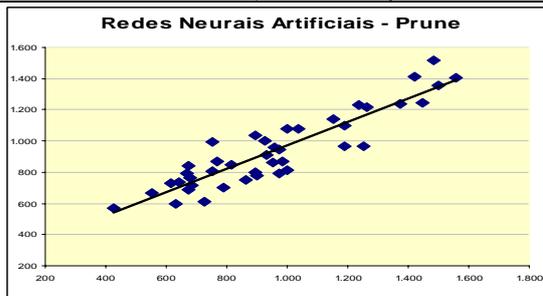


Gráfico 3.18 – Preços x Valores Ajustados

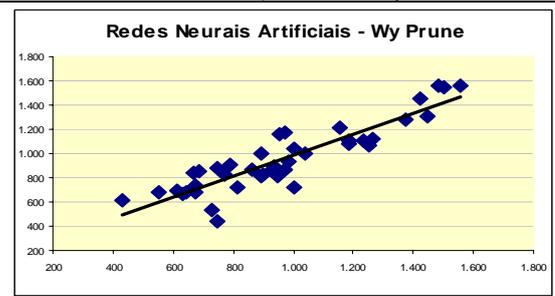


Gráfico 3.19 – Preços x Valores Ajustados

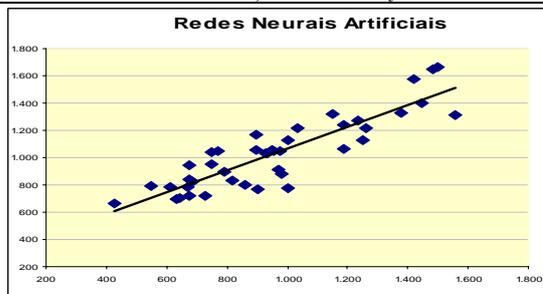


Gráfico 3.20 – Preços x Valores Ajustados

Os gráficos de números 3.16 a 3.20 representam os resultados para a amostra 2.

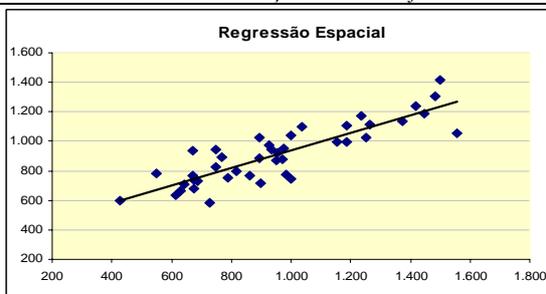


Gráfico 3.21 – Preços x Valores Ajustados

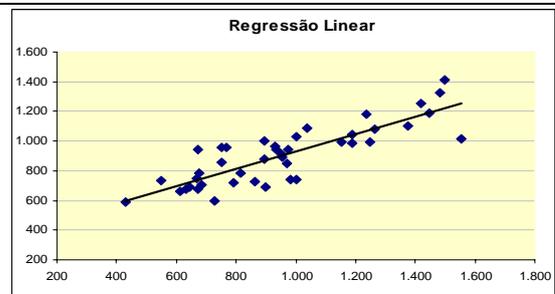


Gráfico 3.22 – Preços x Valores Ajustados

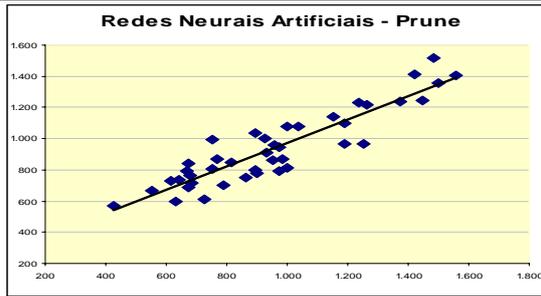


Gráfico 3.23 – Preços x Valores Ajustados

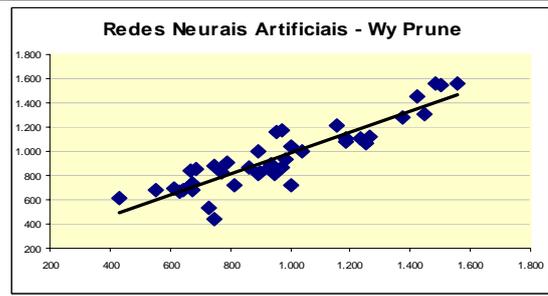


Gráfico 3.24 – Preços x Valores Ajustados

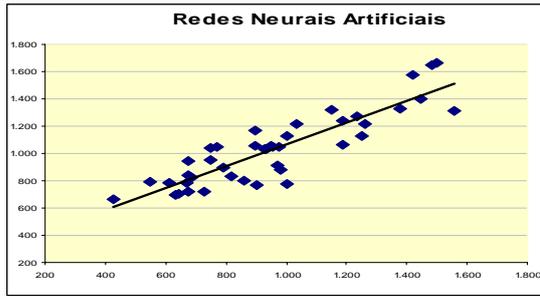


Gráfico 3.25 – Preços x Valores Ajustados

Os gráficos de números 3.21 a 3.25 representam os resultados para a amostra 3.

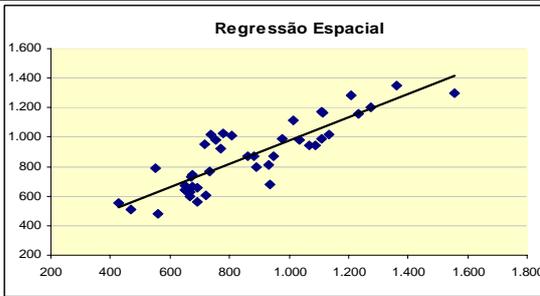


Gráfico 3.26 – Preços x Valores Ajustados

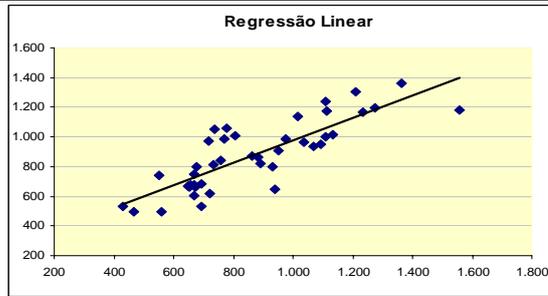


Gráfico 3.27 – Preços x Valores Ajustados

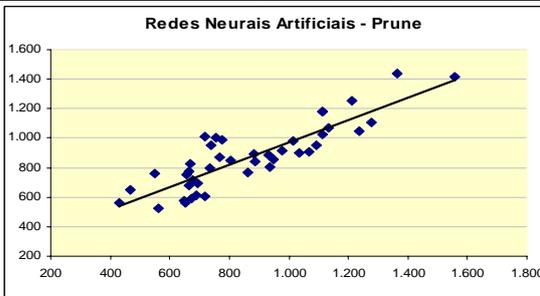


Gráfico 3.28 – Preços x Valores Ajustados

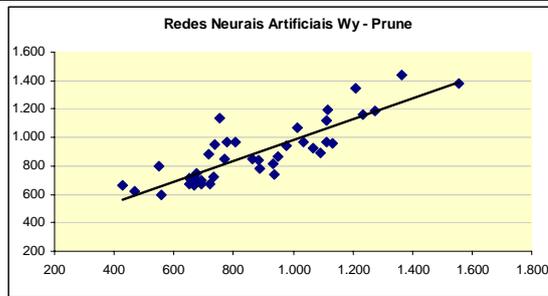


Gráfico 3.29 – Preços x Valores Ajustados

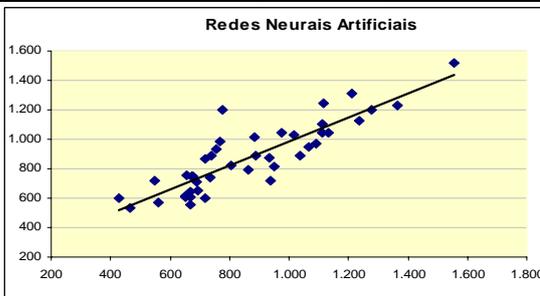


Gráfico 3.30 – Preços x Valores Ajustados

Os gráficos de números 3.26 a 3.30 representam os resultados para a amostra 4.

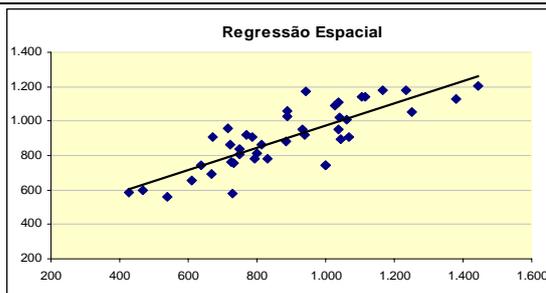


Gráfico 3.31 – Preços x Valores Ajustados

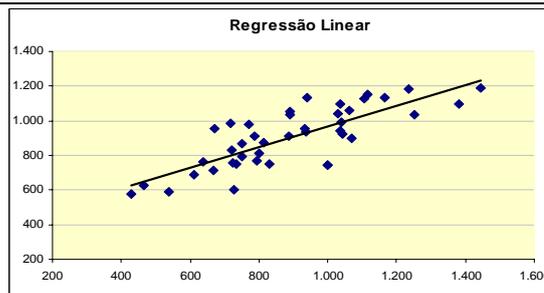


Gráfico 3.32 – Preços x Valores Ajustados

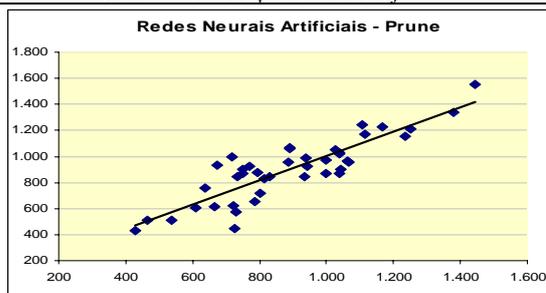


Gráfico 3.33 – Preços x Valores Ajustados

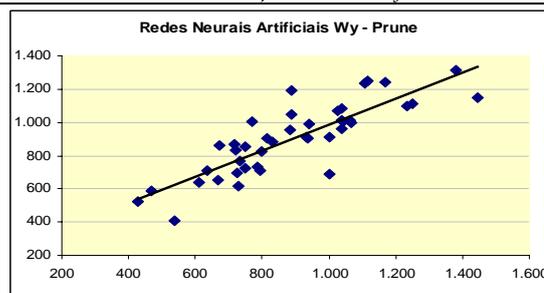


Gráfico 3.34 – Preços x Valores Ajustados

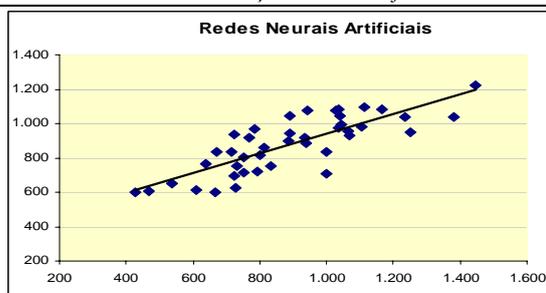


Gráfico 3.35 – Preços x Valores Ajustados

Os gráficos de números 3.31 a 3.35 representam os resultados para a amostra 5.

4. CONCLUSÕES

Pelos resultados encontrados neste trabalho, fica evidenciada a importância da utilização de novas metodologias de modelagem por Redes Neurais Artificiais 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 RNA's tiveram um melhor ajuste tanto na fase de modelagem quanto na 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 a metodologia tradicional não é adequada 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 Redes Neurais Artificiais e, em segundo plano, pelos Modelos de Regressão Espacial, quando a variável de defasagem for incluída em modelos lineares.

Outra questão importante é a forma de considerar a matriz de pesos espaciais na equação de preços hedônicos, que tem sido tomada com base nas distâncias entre os pontos coletados. Esta distância pode ser aferida através de variogramas, mas que infelizmente não detectam divisas físicas (acidentes geográficos) entre os dados. Contudo este problema pode ser solucionado com a introdução das ferramentas de Geoprocessamento, de forma a construir uma matriz de pesos mais eficiente. Ficou demonstrado que existe uma verdadeira interação entre os dados amostrais, de forma que cada edifício funciona com um micro-pólo de

influência sobre os seus vizinhos, mas a introdução da variável em um modelo linear não trouxe resultados melhores do que os obtidos pelas RNA, mesmo sem incluir a matriz de ponderação.

Neste trabalho foram obtidos melhores resultados através das Redes Neurais Artificiais, principalmente quando da aplicação das técnicas de Poda, que proporcionou um melhor poder de generalização à Rede Neural Artificial.

Os resultados obtidos com a Regressão Espacial foram também superiores com relação à regressão linear, contudo inferiores aos modelos de Redes Neurais Artificiais. A introdução da variável de defasagem espacial nos modelos de Redes Neurais Artificiais também não propiciou melhorias nos resultados, indicando que esta variável, para os modelos não lineares, possui baixo poder de explicação, quando tratadas na forma da regressão espacial.

Nas avaliações em massa, principalmente quando houver condições de separar os dados em modelagem e validação, a aplicação das Redes Neurais Artificiais permitirá obter resultados consistentes, quando comparados com a Regressão Linear e a Regressão Espacial, originalmente proposta por Anselin(1998). A Regressão Linear para macro modelos deve ser evitada.

Como sugestão para trabalhos futuros, fica aberto um espaço para profissionais que desejam incorporar, de uma forma mais consistente, a variável de defasagem espacial nos modelos de RNA's. Uma das formas a ser sugerida é o treinamento das redes sem a inclusão da variável de defasagem espacial, para após definida a estrutura da rede, incluir a informação sobre a distribuição espacial dos dados.

BIBLIOGRAFIA

- ANSELIN, L. (1988) “Spatial Econometrics: Methods and Models”. Dordrecht: Kluwer Academic.
- ANSELIN, L. (1988a). “Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity” *Geographical Analysis* 20, 1-17.
- ANSELIN, L. (1990) “SpaceStat: A program for the Statistical Analysis of Spatial Data” (Department of Geography and Department of Economics, University of California, Santa Barbara, CA).
- ANSELIN L. e Rey S. (1991) “Properties of tests for spatial dependence in linear regression models,” *Geographic Analysis*, 23, 112-31.
- ANSELIN, L. e Bera, A. (1998) “Spatial dependence in linear regression models with an introduction to spatial econometrics,” *Handbook of Applied Economic Statistics*, Giles: Marcel Dekker.
- ANSELIN, L. (1999) “Spatial Econometrics”. Bruton Center, School of Social Sciences, University of Texas, Dallas, Richardson, 1999.
- BRAGA, Antonio de Pádua, CARVALHO, André Carlos P. L. Ferreira, LUDEMIR, Teresa Bernarda - – *Redes Neurais Artificiais: Teoria e aplicações* – Rio de Janeiro: LTC, 2000.
- BUSSAB, Wilton de O, MORETTIN, Pedro A - *Estatística Básica* – 5ª Ed. – São Paulo: 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, Pedro Luiz de Oliveira - *Estatística - 18º Reimpressão 2000* – São Paulo: EDGARD BLUCHER, 1977.
- CUN, Y. Le, Boser, B., Denker, J. S., Solla, S. A., *Optimal brain damage: Advances in Neural information processing systems*, vol.2, Morgan Kaufman, San Mateo, 1990, p 598
- DANTAS, R.A. e Cordeiro G.M. (1988) “Uma Nova Metodologia Para Avaliação de Imóveis Utilizando Modelos Lineares Generalizados.” *Revista Brasileira de Estatística*. n. 191, 27-46.
- DANTAS, R. A. (1997) “Avaliação de Imóveis Utilizando Modelos Especiais” IX Congresso Brasileiro de Engenharia de Avaliações e Perícias - São Paulo.
- DANTAS, R. A. (2001) *Engenharia de Avaliações – Uma Introdução à Metodologia Científica* – Ed. PINI - São Paulo.
- DANTAS, R. A. e Cordeiro G.M. (2001) “Evaluation Of The Brazilian City of Recife’s Condominium Market Using Generalized Linear Models” *The Appraisal Journal*, Volume Lxix, Number 3, USA.
- DANTAS, R., Sales, F., Chaves, R. e Magalhães, A. (2001) “Uma nova metodologia para avaliação de imóveis utilizando regressão espacial,” *Anais do XI Congresso Brasileiro de Engenharia de Avaliações e Perícias*, Vitória-ES.
- FAHLMAN, S. E., Libiere, C., *Advances in Neural Information Processing Systems*, Vol. 2, D. S. Touretzky Ed., Morgan Kaufmann, San Mateo, 1990; p 524.
- GONZÁLEZ, Marco Aurélio Stumpf – *Metodologia de Avaliação de Imóveis* – Novo Hamburgo: SGE, 2003.
- GONZÁLEZ, Marco Aurélio Stumpf – *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, Jackson Carvalho – Aplicação de Redes Neurais na Avaliação de Bens - Uma comparação com à análise de regressão - In: II Simpósio Brasileiro de Engenharia de Avaliações (AVALIAR) , 2000 – São Paulo – Anais...São Paulo: AVALIAR, Set.2000 a , p. 3-10.
- GUJARATI, Damodar N. – Econometria Básica – 3ª Ed. – São Paulo: MAKRON, 2000.
- HASSIBI, B., Stork, D. G., In Advances in neural information processing systems 5, Ed.: Morgan Kaufmann, San Mateo, CA, 1993, p 164.
- IBAPE, Diversos Autores (1974) - Anais do I Congresso Brasileiro de Avaliações e Perícias de Engenharia – São Paulo: PINI, 1978.
- KMENTA. Jan –Elementos de econometria – *teoria estatística básica* – V. 1 – 2ª Ed. - São Paulo: ATLAS, 1988.
- KMENTA. Jan – Elementos de econometria – *teoria econométrica básica* – V.2 2ª Ed. – São Paulo: ATLAS, 1988.
- KOVÁCS, Zsolt László - Redes neurais artificiais: fundamentos e aplicações - 3ª Ed. – São Paulo: LIVRARIA DA FÍSICA, 2002.
- HAYKIN, Simon, Redes neurais: princípios e prática - 2º Ed. - Porto Alegre: BOOKMAN, 2001.
- MATHERON, G.(1965) – “*Lés Variables Regionalisées et leur Estimation, Masson*”, Paris.
- MOREIRA, Alberto Lélío – Princípios de engenharia de avaliações - 5a Ed. –São Paulo: PINI, 2001.
- NORMA Técnica da ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS): Avaliação de Bens – Procedimentos Gerais – NBR14653-1. Rio de Janeiro: 2001.
- NORMA Técnica da ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS): Avaliação de Bens – Imóveis Urbanos – NBR14653-2. Rio de Janeiro: 2004.
- 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.
- HAYKIN, S., Neural Networks, a comprehensive foundation, Prentice Hall, Upper Saddle River, New Jersey; 1999, p 218.
- PELLI, Antonio Neto, ZÁRATE, L.E - Avaliação de Imóveis Urbanos com Utilização de Redes Neurais Artificiais - XII Congresso Brasileiro de Engenharia de Avaliações e Perícias – (COBREAP) 2003 – Belo Horizonte . Anais...Belo Horizonte: COBREAP, Set. 2003ª – Resumo - p.14 .
- PELLI, Antonio Neto, Avaliação de Imóveis Urbanos com utilização de Sistemas Nebulosos (Redes Neuro-Fuzzy) e Redes Neurais Artificiais - Congresso Panamericano de Valuación - SCdA UPAV - Cartagena / Colombia - 20 a 23 de setembro de 2004. Trabalho 08 – Versão digital em CD.
- PELLI, Antonio Neto, ZÁRATE, L.E – Valuation of Urban Real Estate Through Artificial Neural Networks - In: Artificial Intelligence and Applications – September 8 -10 ,- Benalmádemá – Spain – A Publication of the International Association of Science and Technology for Development - IASTED – V. 403-169 – p. 523-528 , 2003.
- PELLI, Antonio Neto, BRAGA, A. P. – Redes Neurais Artificiais : aplicação e comparação dos resultados com regressão linear na avaliação de imóveis urbanos – IV concurso Internacional de Valuación y Catastro – SOITAVE – Caracas/ Venezuela – 17 a 20 de Abril de 2005 – versão digital em CD.

- PUCCINI, Abelardo de Lima – Matemática financeira – Objetiva e Aplicada - 6ª Ed. – São Paulo: SARAIVA 2002.
- RUSSELL, Stuart J., NORVING, P. – Inteligência Artificial - 2ª Ed. – Rio de Janeiro: ELSEVIER, 2004.
- STEWART, James – Cálculo – Volume I – 4ª Edição - São Paulo: PIONEIRA , 2003;
- WERKEMA, Maria Cristina Catarino - AGUIAR, Silvio - 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.

CURRICULUM VITAE

Nome: Antônio Pelli Neto

Empresa: Caixa Economica Federal **Cargo:** Engenheiro **Função:** Assistente Técnico I

Endereço Comercial: Av. João Pinheiro, 584 - Bairro Funcionários – BH/MG.

Telefones: +55 (31) 3248-5200 – 3248-5229 - **Fax:** +55 (31) 3248-5200

Graduação: Engenharia Civil e Mecânica UFMG - Universidade Federal de Minas Gerais – 1982 - **CREA:** 6021/D-DF

Mestrado: Engenharia Elétrica, Inteligência Computacional - Universidade Federal de Minas Gerais – UFMG – 2006

E-mail: antonio-pelli.neto@caixa.gov.br - pelli@terra.com.br

PRINCIPAIS PARTICIPAÇÕES EM CONGRESSOS E EVENTOS

- Apresentação de trabalho: **Informatizacion de Evaluacion Econômica de Empreendimento** no Iº Congresso Internacional Valoración Y Tasacion – Valência - Espanha – 03 a 05 de Julho de 2002;
- Apresentação de trabalho: **Avaliação de Imóveis Urbanos com utilização das Redes Neurais** no XII COBREAP – Congresso Brasileiro de Engenharia de Avaliações e Perícias – Belo Horizonte/MG – 02 a 05 de Setembro de 2003. *Agraciado com a Menção Honrosa Orlando Andrade Resende concedida pelo Instituto Brasileiro de Avaliações e Perícias de Engenharia – IBAPE-MG;*
- Apresentação de trabalho: **Artificial Intelligence and Application's** – IASTED – International Association of Science And Technology dor Development - Benalmadema, Espanha, 08 a 10 September de 2003;
- Apresentação do Trabalho: **Avaliação de Imóveis Urbanos com utilização de Sistemas Nebulosos (Redes Neuro-Fuzzy) e Redes Neurais Artificiais**, no XXI Congresso Panamericano Valuación – UPAV - Cartagena/Colômbia - 20 a 23 Setembro 2004;
- Apresentação do Trabalho - **Redes Neurais Artificiais: aplicação e comparação dos resultados com a regressão linear nas avaliações de imóveis urbanos**, durante o X Congresso Venezolano de Valiación Y Catastro e I Concurso Internacional de Valuación e Catastro , dias 17 a 20 de Abril de 2005. *Agraciado com o 1º lugar no IV Concurso Internacional de Valuación Y Catastro Caracas-VE;*

ATIVIDADES PROFISSIONAIS - Caixa Econômica Federal (a partir de 1982)

- Divisão de Processamento de Dados.
- Desenvolvimento de Programas - Analista de Sistemas.
- Concurso Público - Engenheiro Mecânico e Civil.
- Participação do Grupo de Trabalho de Avaliação dos Imóveis Funcionais da União.
- Desenvolvimento do *Sistema de apoio a Avaliação Imobiliária – SAAI*
- Desenvolvimento do *Sistema de Avaliações Imobiliárias – SIM.*

Designado Instrutor de Avaliação de Imóveis – Avaliar Bens, tendo ministrado diversos cursos de Avaliação de Imóveis no âmbito da Caixa Econômica Federal.