

# Modelo de Regressão Linear Múltipla para a Predição do Preço de Imóveis Residenciais Urbanos no Município de Paranaíba-MS.

*A Multiple Regression Linear Model to Predict Residential Real Estate  
Price in the Municipality of Paranaíba-MS*

Elton Gean Araújo\*

Júlio César Pereira†

Vinicius Renan Araújo‡

João Paulo RISSO§

Ricardo André Hornburg¶

**Resumo:** O presente artigo aplica a análise de regressão linear múltipla a dados de imóveis residenciais, para previsão do preço do imóvel a venda no município de Paranaíba-MS em função de suas características (covariáveis). A metodologia foi aplicada a um conjunto de 45 casas que estavam à venda nos sites imobiliários do município. O método de seleção de variáveis *Stepwise* foi aplicado, em que as covariáveis Localização (Centro ou fora do Centro), Área do Terreno, Número de Salas, Estado de Conservação (Baixa, Média e Alta), Vagas de Garagem foram selecionadas para comporem o modelo. O modelo final apresentou um ajuste adequado aos dados e uma capacidade de predição bastante satisfatória, tornando-se assim uma ferramenta adicional confiável para avaliação de imóveis urbanos no município.

**Palavras-chave:** Modelagem. Predição. Mercado imobiliário.

**Abstract:** In this paper we aimed to apply multiple linear regression analysis to evaluate residential properties for sale in the city of Paranaíba-MS, Brazil, according to their characteristics (covariates). The methodology was applied to a data set composed by 45 houses

---

\*Doutor, Universidade Federal de Mato Grosso do Sul, [egarauj@yahoo.com.br](mailto:egarauj@yahoo.com.br)

†Doutor, Universidade Federal de São Carlos, [julio-pereira@ufscar.br](mailto:julio-pereira@ufscar.br)

‡Cursando Licenciatura em Matemática, Universidade Federal de Mato Grosso do Sul, [viaraujo@hotmail.com](mailto:viaraujo@hotmail.com)

§Cursando Licenciatura em Matemática, Universidade Federal de Mato Grosso do Sul [joaor-mat@gmail.com](mailto:joaor-mat@gmail.com)

¶Doutor, Faculdade Avantis, [ricardo\\_andreh@yahoo.com.br](mailto:ricardo_andreh@yahoo.com.br)



that were for sale in the state agents websites of the municipality. For variable selection stepwise method was applied. The coariates Location (Center or Outside Center), Land Area, Number of Rooms, State of Conservation (Low, Medium and High), Garage were selected to compose the model. The final model presented a suitable adjustment to the data and a very satisfactory prediction capacity. In conclusion the proposed method is an additional and reliable tool for the evaluation of urban properties in the municipality.

**Keywords:** Modelling. Prediction. Real estate market.

## 1 Introdução

O município de Paranaíba está localizado nas fronteiras dos estados de São Paulo, Minas Gerais e Goiás, mais especificamente situa-se no estado de Mato Grosso do Sul e na região Centro-Oeste do Brasil, tendo como principais atividades econômicas a pecuária, agricultura e, por compor a Região do Bolsão Sul-Mato-Grossense, apresenta potencialidades de negócios locais, em especial, aquelas voltadas para as microempresas e empresas de pequeno porte. Além disso, outro aspecto importante condiz com a presença de universidades públicas (estadual e federal) e particulares, o que tem sido um fator importante para a expansão do município [14].

De acordo com o último censo realizado pelo Instituto Brasileiro de Geografia e Estatística [10], a população de Paranaíba-MS, no último censo, realizado em 2010, era de 40.192 pessoas, em 2016, foi estimada em 41.626 habitantes. Localiza-se a uma latitude de  $19^{\circ}40'38''$  Sul e a uma longitude  $51^{\circ}11'27''$  Oeste. Sua área territorial é de  $5.402,652 \text{ km}^2$ .

O setor imobiliário está sendo impulsionado pelo desenvolvimento de novos bairros (residenciais Califórnia e Constroluz); aumento do PIB per capita – segundo dados do IBGE de 2010 e 2014, tal índice cresceu de R\$ 15.220,51, em 2010, para R\$ 24.464,35 em 2014; crescimento populacional: entre os anos censitários de 2000 e 2010, a quantidade de habitantes do município de Paranaíba aumentou 5% e o número de domicílios cresceu 20% no mesmo período, passando de 11.399 para 13.700 domicílios no município [10].

No contexto da avaliação de imóveis, percebe-se a necessidade de desenvolvimento de metodologias estatísticas (caracterizadas pela precisão científica) de precificação de imóveis residenciais. Além deste aspecto, após uma vasta busca em periódicos pelo *Web of Science* e Google Acadêmico, notou-se o raro uso de tais metodologias em artigos acadêmicos e conseqüentemente por profissionais do mercado imobiliário.

Face ao exposto, o objetivo deste trabalho é propor um modelo de regressão linear múltipla para predição do preço de casas à venda no município de Paranaíba-

MS, que se apresenta como metodologia científica de avaliação de imóveis, a partir de suas características físicas e geográficas.

## 2 Referencial teórico

No âmbito das metodologias de avaliação de imóveis, a modelagem estatística é fundamental, pois apresenta-se como meio sistemático e facilitador na determinação de inferências de fenômenos a partir de amostras do todo e, segundo [5],

[...] inferir significa concluir. Assim, inferir estatisticamente significa tirar conclusões com base em medidas estatísticas. Em Engenharia de Avaliações o que se pretende é explicar o comportamento do mercado que se analisa, com base em alguns dados levantados no mesmo. Neste caso a inferência estatística é fundamental para solucionar a questão, pois conhecendo-se apenas uma parte do mercado pode-se concluir sobre o seu comportamento, com determinado grau de confiança. [5] ( p.69)

Segundo [16] , as metodologias, normatizadas (determinadas pela ABNT, por exemplo), que possibilitam a realização de avaliações de imóveis, baseiam-se em três métodos básicos: métodos para identificarem o valor de um imóvel; métodos para identificarem o custo de um imóvel e, métodos para levantar indicadores de viabilidade da utilização econômica do imóvel. Além disso, [16] aponta que devido os imóveis apresentarem características heterogêneas, é necessária uma análise das mesmas optando-se por uma das seguintes metodologias: análise por fatores ou análise científica. O método comparativo direto de dados de mercado é o mais utilizado na avaliação de imóveis urbanos [19]. Em tal método, o valor do imóvel é aferido por meio da comparação com dados congêneres do mercado [5].

No âmbito da modelagem estatística, [2] propuseram um modelo de regressão linear múltipla (modelo de regressão linear que possui mais de uma variável independente incorporada) como metodologia de avaliação do preço de venda de imóveis residenciais em Bonito/MS. Esse trabalho objetivava, através de uma abordagem quantitativa, identificar características que contribuem para explicar o preço, bem como encontrar um modelo para estimar o preço de imóveis em tal município. Com objetivos semelhantes, [13] também desenvolveram um modelo de regressão linear múltipla para estimar preços de casas a venda em Sorocaba-SP. Ambos os trabalhos se preocuparam com a seleção de variáveis que realmente pudessem contribuir com a construção dos modelos citados, com o intuito de fomentar a veracidade das avaliações dos imóveis à venda.

Atualmente, estudos e construções de modelos que consideram a variável localização e a relação que os imóveis estabelecem entre si, a partir da mesma, vem

contribuindo à realização de eficientes e rigorosas avaliações de imóveis à venda. Neste sentido, pode-se apontar o recente trabalho de [9] realizado em Balneário Camboriú-SC, na qual apresenta um método que relaciona o uso da econometria espacial com a geoestatística bayesiana, tendo em vista estimar o valor dos imóveis levando-se em consideração as interações espaciais relevantes às características da localização.

Outro relevante trabalho, que se caracterizou pela análise espacial, foi realizado por [20], na qual a coleta de dados para aplicação ocorreu no município de São José-SC. Segundo os autores, o método proposto possibilitou incluir a localização geográfica o qual foi vigorosamente significativo no modelo de regressão para todos os tipos de imóveis da região e, permitiu modelar de modo coeso o efeito do conjunto de polos de valorização na região. [20] apontam que tal método exclui toda autocorrelação espacial nos resíduos do modelo de regressão, aprimorando o poder explicativo e a confiabilidade da avaliação.

### 3 Procedimentos metodológicos

Os dados utilizados nesta pesquisa compõem uma amostra de casas no município de Paranaíba-MS, que estavam à venda em imobiliárias da cidade. A amostra é formada por todas as 51 casas disponíveis para venda nos dias de coleta, sendo que 45 delas foram selecionadas aleatoriamente para ajuste do modelo e 6 foram deixadas de fora para posterior validação do modelo. A amostra das casas não contempla todos, mas contempla diversos bairros da cidade. O período de coleta ocorreu no mês de maio de 2017.

Outras características relacionadas que compõem o conjunto de dados são: localização, Latitude, Longitude, quantidade de quartos (NQ), vagas de garagem (VG), área total do terreno (AT), número de banheiros (NB), salas (NSA), piscina (P), estado de conservação (Cons) e preço. Neste caso, a variável resposta  $Y$  refere-se aos preços dos imóveis e as outras variáveis são consideradas independentes. A variável localização (Centro) foi classificada como qualitativa nominal dicotômica (*dummy*), onde se atribuiu o código 0 para as casas fora do centro e 1 para as casas situadas no centro. A amostra constituiu-se de um conjunto de 45 imóveis residenciais urbanos que estavam à venda nos sites imobiliários do município.

O método de pesquisa utilizado trata-se de pesquisa quantitativa baseada em modelo, pois se procurou construir, analisar e testar um modelo que estabelece relações de causa e efeito entre variáveis. Assumiu-se uma relação de causa e efeito entre preço de imóveis e as demais variáveis coletadas. A partir disso, buscou-se

criar um modelo que descrevesse adequadamente essas relações, permitindo assim o entendimento do comportamento da variável preço [3] [2].

Considerando-se  $k$  regressores (covariáveis) o modelo de regressão linear múltipla pode ser expresso pela equação

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon, \quad (1)$$

em que  $\beta_j$ ,  $j = 0, 1, \dots, k$  são os coeficientes de regressão, sendo parâmetros que representam a variação esperada em  $y$  por unidade de variação em  $X_j$ , quando todos os outros regressores são mantidos constantes ([12]).

A seleção de covariáveis para fazerem parte do modelo foi feita através do procedimento *Stepwise* (passo a passo), nas direções "*forward*", "*backward*" e "*both*" e avaliadas em cada etapa por um critério de comparação [13]. No método passo a frente (*forward*), as covariáveis devem ser adicionadas ao modelo, uma de cada vez, até que não haja mais covariáveis que produzam um modelo com melhor poder preditivo do que o modelo obtido previamente, sendo que o processo inicia-se sem nenhuma variável; no método passo atrás (*backward*), o modelo se inicia com todas as covariáveis, sendo retiradas, uma de cada vez, e, excluídas do modelo as variáveis que não contribuam para um incremento significativo na explicação do preço dos imóveis; o método passo à frente e passo atrás (*both*) combina as duas direções anteriores.

O critério utilizado para eliminar ou não uma covariável do modelo nos métodos de seleção *Stepwise* foi o AIC (Critério de Informação de Akaike), mais indicado para pequenas amostras, como é o caso; e, segundo [11], o melhor modelo é o que apresenta menor AIC. O método de estimação dos parâmetros do modelo de regressão linear múltipla utilizado neste trabalho foi o método dos mínimos quadrados.

Após a seleção das covariáveis que fariam parte do modelo, foi avaliado o resíduo do modelo com o intuito de verificar se os mesmos não violavam as pressuposições de um modelo de regressão linear, tal como normalidade e independência.

Devido ao fato de se ter observações referenciadas espacialmente, por coordenadas geográficas é comum o resíduo do modelo de regressão linear apresentar dependência espacial. Nesse caso, espera-se que casas que se encontram localizadas em pontos próximos entre si apresentem maior semelhança no preço (correlação espacial) enquanto que, para casas mais distantes umas das outras, espera-se que os preços sejam mais independentes [17]. Por esse motivo, fez-se uma análise espacial exploratória a fim de se detectar possível correlação.

Uma forma de explorar as relações que possam existir entre unidades amostrais espacialmente referenciadas é fazendo o uso da geoestatística [8]. Como ferramenta

para a detecção de correlação espacial, foi utilizado o semi-variograma empírico. Em poucas palavras o semivariograma empírico é o gráfico da semivariância observada em função da distância. A correlação espacial é detectada quando se observa um aumento na semivariância com o aumento da distância. Detalhes sobre essa ferramenta podem ser encontrados em [8].

Além da análise espacial exploratória, ajustou-se também um modelo de regressão linear múltipla adicionado de uma componente aleatória para acomodar possível correlação espacial nos dados. Tal modelo foi então comparado com o modelo obtido previamente no método de seleção de variáveis (eq. 1).

O modelo de modelo de regressão linear múltipla adicionado de uma componente aleatória para acomodar possível dependência espacial é descrito pela equação

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + v + \epsilon \quad (2)$$

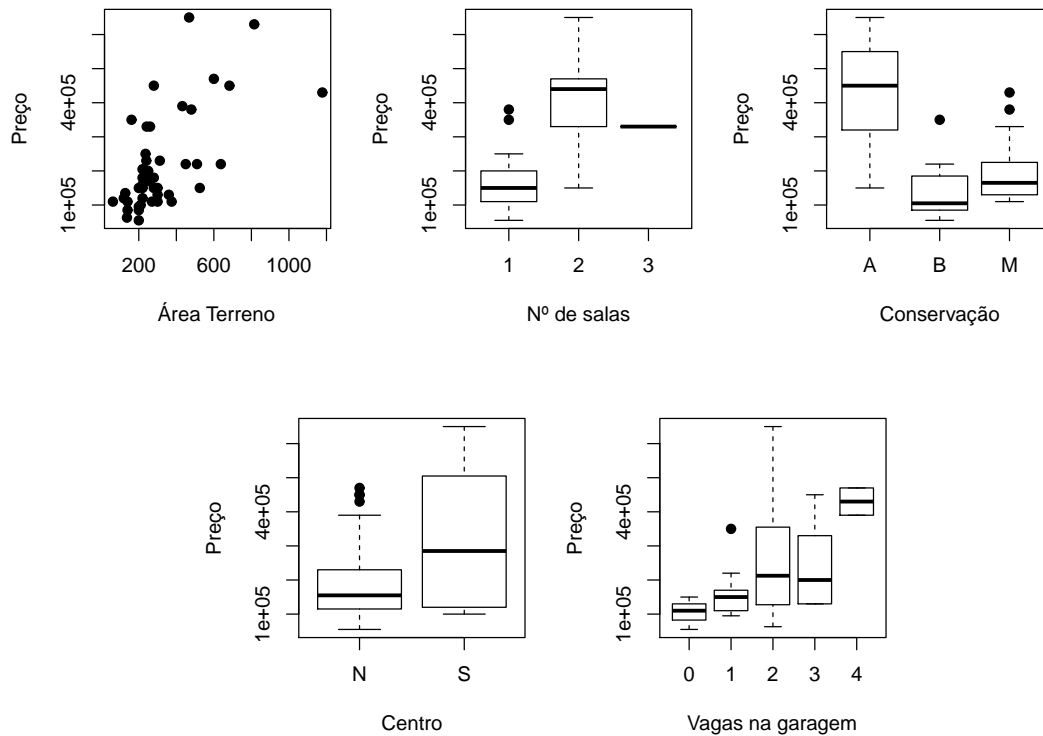
em que  $v$  é uma componente aleatória com distribuição normal multivariada, com vetor de médias zero e matriz de covariâncias  $\Sigma$ , tal que  $\Sigma_{ij} = \sigma^2 \rho(d, \phi)$ , sendo  $\sigma^2$  a variância do processo espacial e  $\rho(d, \phi)$  representa a função de correlação que depende do parâmetro  $\phi$  e das distâncias  $d$ . A função de correlação adotada foi a exponencial.  $\epsilon(s)$  é uma componente de ruído branco, normalmente distribuída, com média 0 e variância  $\tau^2$  (efeito pepita), que descreve a variação de microescala.

Toda a análise estatística, ajuste de modelos, seleção de variáveis e cálculo de critérios de comparação de modelos foram realizadas com o auxílio do software R [15].

## 4 Resultados

Por meio do método de seleção de variáveis, citado na seção Procedimentos Metodológicos, o modelo selecionado apresentou o menor valor de AIC (999,86) e os mesmos resultados para as direções *backward*, *forward* e *both*. As variáveis selecionadas por estes métodos foram Área do Terreno, Número de Salas, Localização (Centro ou fora do Centro), Estado de Conservação e Vagas de Garagem.

Na Figura 1, pode ser observado o comportamento da variável resposta Preço, com as demais covariáveis selecionadas pelo modelo de regressão linear múltipla. Pode-se notar grande associação entre as variáveis Preço e as covariáveis. Isso mostra que, todas as covariáveis selecionadas pelo modelo têm forte influência (positiva) na composição dos preços dos imóveis no município de Paranaíba – MS. Este fato pode ser comprovado pelo coeficiente de correlação de Pearson entre as covariáveis quantitativas que compõe o modelo e a variável resposta Preço (Tabela 1).



**Figura 1:** Gráficos da Variável resposta preço vs as covariáveis selecionadas pelo modelo.

**Tabela 1:** Coeficiente de Correlação de Pearson entre as covariáveis quantitativas do modelo e a variável resposta Preço dos imóveis.

	Área Terreno	N° de Salas	Vagas de Garagem
Valor	0,65*	0,89*	0,83*
Área Terreno		0,55*	0,42*
N° de Salas			0,85*

\* p-valores menores que 0,05

No que tange às covariáveis qualitativas do modelo, pode-se observar que as casas com estado de conservação Baixa, Média e Alta, tem valores medianos de preço respectivamente crescentes nessa ordem (Figura 1), do mesmo modo, mostra-se um incremento no preço dos imóveis se as casas estiverem localizadas no centro.

O modelo selecionado,

$$\begin{aligned}
 \text{Valor} = & 78614,2 + 115669,4 \text{Centro} + 20105,4 \text{VG} + 190,3 \text{AT} \\
 & + 105992,1 \text{NSA} - 178893,3 \text{ConsB} - 113696,3 \text{ConsM},
 \end{aligned} \quad (3)$$

mostra a magnitude dos efeitos das covariáveis selecionadas sobre o preço dos imóveis por meio dos coeficientes de regressão.

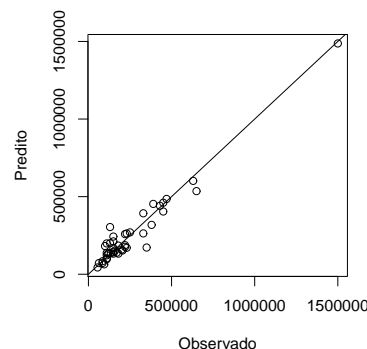
A Tabela 2 apresenta a análise de variância do modelo de regressão linear múltipla selecionado, mostrando a porção da variabilidade explicada por cada covariável, bem como os valores das estatísticas dos testes de significância e seus respectivos p-valores para cada uma delas.

**Tabela 2:** Análise de Variância do modelo selecionado.

Covariável	GL	SQ	QM	F	p-valor
AT	1	1,08E+12	1,08E+16	27,55	$<2,2e-16^{***}$
NSA	1	1,02E+16	1,02E+16	264,21	$<2,2e-16^{***}$
Cons	2	9,97E+14	4,99E+14	12,88	$5,366e-05^{***}$
Centro	1	7,72E+14	7,72E+14	19,94	$6,949e-05^{***}$
VG	1	9,15E+14	9,15E+14	23,63	$2,049e-05^{***}$
Resíduo	38	1,47E+15	3,87E+13		

\*\*\* p-valores menos que 0,001 do teste de significância dos parâmetros do modelo

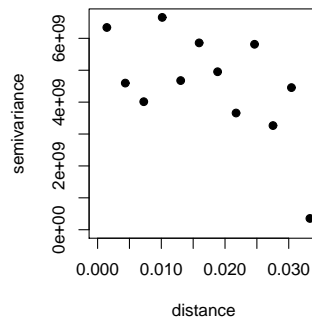
A Tabela 2 mostra que todas as covariáveis foram significativas pelo teste F, em que todas apresentaram p-valores menores que 0,05. Tal fato mostra que todas as covariáveis selecionadas são muito importantes na predição do preço de imóveis residenciais no município de Paranaíba-MS. O coeficiente de determinação ajustado foi igual a 0,94, isso mostra que 94% da variação dos preços de imóveis foi explicada pelo modelo selecionado. O gráfico dos preços preditos das imóveis vs preços observados dos imóveis residenciais é apresentado na Figura 2.



**Figura 2:** Gráfico dos preços observados dos imóveis *vs* preços preditos dos imóveis residenciais a venda no município de Paranaíba-MS.

Na Figura 3 é apresentado o semivariograma empírico construído a partir do resíduo do modelo ajustado (Eq. 3).



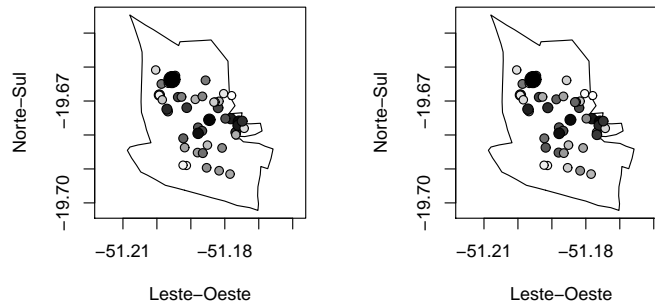


**Figura 3:** Semivariograma empírico para o resíduo do modelo da Eq. 3.

Pelo variograma não se observa qualquer sinal de dependência espacial do resíduo, pois não se observa uma tendência na semivariância com o aumento da distância. Apesar do não se ter detectada nenhuma dependência espacial no resíduo do modelo da eq. (3), optou-se mesmo assim por comparar o ajuste do modelo da eq. (3), com sua versão adicionada de uma componente espacial, modelo este representado teoricamente pela eq. (2). O valor de AIC obtido para o modelo espacial foi de 1005 contra 999,86 do modelo não espacial. Ou seja, o modelo espacial apresentou maior AIC do que o modelo não espacial. Esse resultado corrobora o fato de não se ter detectado correlação espacial pela análise do variograma. Pelo fato de não se ter correlação espacial nos dados o modelo espacial acaba penalizado pelo aumento no número de parâmetros, isso também faz com que apresente maior valor de AIC e portanto, seu uso não se torna viável nesta situação. Dessa forma, o modelo apresentado na eq. (3) é preferível para descrever e fazer inferências sobre o preço de imóveis nessa situação. Mesmo não tendo sido observada dependência espacial na distribuição dos preços dos imóveis, optou-se mesmo assim por construir um mapa (Figura 4) contendo a distribuição dos preços dos imóveis de acordo com o mapa da cidade de Paranaíba-MS, o que de certa forma contribuiu para uma melhor visualização dos preços dos imóveis. Os mapas apresentados mostram os preços observados e os preços preditos pelo modelo da eq. (2). Nota-se grande semelhança entre os dois mapas, retratando a acurácia do modelo nas predições.

#### 4.1 Diagnóstico e validação do modelo

Apesar das variáveis explicativas quantitativas do modelo selecionado estarem significativamente correlacionadas (Tabela 1), tal fato não justifica a multicolinearidade (Tabela 3), tampouco prejuízo nas estimativas dos coeficientes de regressão, pois segundo [7], valores de VIF maiores que 10 podem causar problema nas estimativas, que não é o caso deste trabalho.



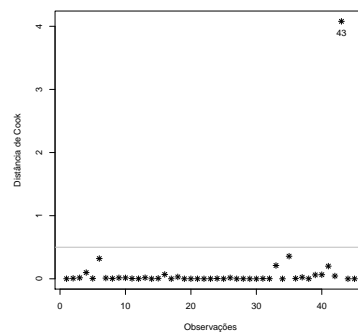
**Figura 4:** Preço observado à esquerda e preço predito à direita sobre o mapa da cidade de Paranaíba. O tamanho do círculo é proporcional ao preço.

**Tabela 3:** Diagnóstico de multicolinearidade com base no fator de inflação da variância.

Variáveis	AT	NSA	VG
VIF	1,57	4,74	3,60

VIF é a abreviação de variance inflation factors

Na Figura 5, tem-se as distâncias de Cook's, que avaliam observações influentes no modelo ajustado. De acordo com [4], observações com distâncias maiores que 1, podem ser consideradas influentes para o modelo. Nota-se pela Figura 1, que apenas a observação 43 pôde ser considerada influente; porém, foi realizada nova análise sem tal observação, em que não houve mudança nas significâncias dos coeficientes de regressão, bem como acréscimo em apenas um centésimo no coeficiente de determinação, portanto, a observação foi mantida no modelo.



**Figura 5:** Gráfico com os valores das distâncias de Cook's.

Com exceção da avaliação dos outliers, todos os outros testes apontaram para o

atendimento das pressuposições do modelo de regressão selecionado (Tabela 4). O teste de Outlier de Bonferroni apontou dois candidatos a outlier, as observações 6 e 35. Considerando que os p-valores ficaram próximos do nível de significância 0,05 (0,040 e 0,042, respectivamente), optou-se por manter as observações.

**Tabela 4:** Resultados dos testes estatísticos de diagnóstico do modelo de regressão selecionado.

Testes Diagnósticos	ET	RE	p-valor
Teste Outlier de Bonferroni		3,640	0,040
Teste Outlier de Bonferroni		-3,590	0,042
Teste de Homocedasticidade - Goldfeld-Quandt		1,42	0,250
Teste de Normalidade - Shapiro Wilk	0,96		0,080
Teste de Independência - Durbin Watson	2,03		0,544

RE = Resíduos estudentizados; ET = Estatística do Teste

Por fim, para efeitos de validação do modelo de regressão selecionado, coletou-se adicionalmente uma amostra de 6 (seis) imóveis residenciais à venda em imobiliárias do município. Na Tabela 5 verificam-se os erros absolutos e percentuais em relação aos valores observados das casas; nota-se que os erros percentuais variaram entre 1,12 e 5,78 %, o que indica um modelo com boa capacidade preditiva.

**Tabela 5:** Resultados da Validação do Modelo Selecionado.

Valor Obs.	Valor Predito	Erro Absoluto	Erro (%)
140000	148098,90	8098,90	5,78
80000	82901,85	2901,85	3,63
140000	141565,33	1565,33	1,12
75000	73387,94	1612,06	2,15
300000	283710,27	16289,73	5,43
105000	108965,68	3965,68	3,78

## 5 Discussão

De acordo com os resultados da análise exploratória dos dados (Figura 1 e Tabela 1), pode-se observar forte associação entre as covariáveis que compuseram o modelo (selecionadas pelo método *Stepwise*) e a variável resposta preço dos imóveis. Além disso, o teste F da análise de variância (Tabela 2), mostrou que todas as covariáveis selecionadas (AT, NSA, Cons, Centro e VG) foram significativas ao nível de 5% de

significância, o que permite concluir que todas estas covariáveis têm forte influência na composição do preço dos imóveis.

Em trabalhos que tem utilizado a técnica de regressão linear múltipla para avaliação do preço de imóveis residenciais urbanos, como o de [2] e [13], obtiveram que o modelo selecionado foi composto por área útil, área do terreno, vagas de garagem, número de dormitórios e localização; neste sentido, as covariáveis selecionadas para compor o modelo neste trabalho e que influenciam na explicação dos preços dos imóveis corroboram com os resultados obtidos pelos referidos trabalhos.

Os resultados da Tabela 2 e o coeficiente de determinação ajustado (0,94) indicam que todas as covariáveis selecionadas são indispensáveis para o modelo, além disso 94% de toda variação no preço dos imóveis é explicada pelo modelo. Por fim, a Figura 2, mostra que o modelo tem uma boa capacidade preditiva, pois os preços observados dos imóveis *vs* os preços preditos dos imóveis são bem ajustados pela reta de regressão.

Para utilização do modelo, devem-se substituir os valores das covariáveis em (2) pela área do terreno (AT), número de salas (NSA), estado de conservação (Cons), localização (Centro) e vagas de garagem (VG) da casa que se deseja estimar o preço de venda. Nota-se pelo modelo da eq. (2) que cada  $m^2$  a mais na área do terreno, há um acréscimo de R\$ 190,3 no preço do imóvel, mantendo-se constantes as demais covariáveis. Com relação ao número de salas, a cada sala a mais no imóvel, seu preço aumenta R\$ 105992,10; a cada vaga de garagem a mais no imóvel, o preço sofre um acréscimo de R\$ 20105,40; e, para uma casa localizada no centro da cidade, seu valor é R\$ 115669,40 maior do que uma casa não localizada no centro da cidade, considerando-se todas as outras covariáveis fixas.

Por fim, se o estado de conservação do imóvel for considerado "Médio" há uma desvalorização de R\$ 113696,30 comparado a um imóvel nas mesmas condições, porém com estado de conservação considerado "Alto". Quando o estado de conservação é considerado "Baixo" a desvalorização é de R\$ 178893,30 em relação a um imóvel com estado de conservação "Alto".

Em [2], quando propuseram um modelo de regressão linear múltipla para avaliação de imóveis residenciais no município de Bonito-MS, a covariável Local (centro ou fora do centro) foi a que mais influenciou no preço dos imóveis, sendo que, uma casa no centro tem um valor acrescido de R\$ 88600,90, quando comparado a um imóvel fora do centro, nas mesmas condições; e, a cada vaga de garagem adicional no imóvel, o acréscimo esperado foi de R\$ 32469,90. Para [13], em que se construiu um modelo de regressão linear múltipla para predição dos preços de imóveis no município de Sorocaba-SP, a área do terreno impactou de forma mais acentuada o preço

do imóvel do que neste trabalho, sendo que, o aumento de cada  $m^2$  proporcionou um acréscimo esperado de R\$ 451,90, mantendo-se as demais covariáveis constantes no modelo.

Em [6], quando estudaram a importância dos vizinhos na avaliação de imóveis no caso Recife-PE, concluíram que para cada nível de melhor conservação do imóvel, este sofre uma valorização de 7,36%, e, quando incorporado a componente espacial no modelo, houve uma desvalorização de 6,87% no preço, para cada nível de pior conservação do imóvel. Em [18], aplicaram-se modelos estatísticos multivariados na avaliação de imóveis em Campo Mourão-PR, a covariável número de salas apresentou peso fatorial aproximado -0,36 para a terceira componente principal, a qual está relacionada ao número de cômodos da casa; posteriormente esta componente tomada como covariável no modelo de regressão linear múltipla apresentou significância estatística ao nível de 5%, evidenciando que o número de salas contribuiu na formação dos preços dos imóveis, como neste trabalho.

## 6 Considerações finais

Neste trabalho foi proposta uma metodologia baseada em modelos de regressão linear múltipla para a predição de preços de imóveis residenciais a venda na cidade de Paranaíba-MS. O método utilizado para seleção das variáveis do modelo foi o *Stepwise*.

A metodologia apresentada para realização de predição de preços de casas se mostrou viável e adequada, com poder preditivo satisfatório. Todas covariáveis selecionadas no modelo final, área do terreno, vagas de garagem, localização, número de salas e estado de conservação do imóvel foram estatisticamente significativas para o modelo, sendo importantes para a composição dos preços dos imóveis no município de Paranaíba-MS.

Assim, o modelo proposto pode servir de apoio tanto a gerentes imobiliários na definição de preços de casas, como para as pessoas físicas e/ou jurídicas que desejam avaliar seu patrimônio de forma genuína. Deve-se atentar que os modelos de regressão linear múltipla definidos, deverão sofrer reajustes periódicos, ou seja, deverão ser atualizados, devido a alta dinamicidade da economia do país.

## Referências

- [1] ARAUJO, E.G.; PEREIRA, J.C.; ARAÚJO, V.R.; RISSO, J.P., HORNBERG, R.A. Modelo Multivariado para Predição do Preço de Imóveis Residenciais Ur-



banos no Município de Paranaíba-MS. **I SIMSAD - UFMS**, 2017.

- [2] ARAUJO, E.G.;PEREIRA, J.C.; XIMENES, F.; SPANHOL, C.P.; GARRISON,S. Proposta de uma metodologia para a avaliação do preço de venda de imóveis residenciais em Bonito/MS baseado em modelos de regressão linear múltipla. **Pesquisa & Desenvolvimento Engenharia de Produção**, v. 10, p. 195-207, 2012.
- [3] BERTRAND, J.W.M.; FRANSOO, J.C. Modelling and simulation: Operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management.**, v. 22, n.2, p. 241-264, 2002.
- [4] COOK, R.D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman and Hall, 1982.
- [5] DANTAS, R.A. **Engenharia de Avaliações: uma introdução à metodologia científica**. São Paulo: Pini, 2005, 255 p.
- [6] DANTAS, R.A.; MAGALHÃES, A.M.; VERGOLINO, J.R.O. Avaliação de imóveis: a importância dos vizinhos no caso de Recife. **Economia Alicada**, São Paulo, v. 11, n. 2, p.231-251, 2007.
- [7] DRAPER, N. R. ; SMITH, H. **Applied Regression Analysis**. 3a. edição. John Wiley and Sons, EUA, 706 p, 1998.
- [8] DIGGLE, P.;RIBEIRO JR, P.J.**Model-based Geostatistics**.. Series: Springer Series in Statistics, 2007, 230 p.
- [9] HORNBURG, R.A.; HOCHHEIM, N.Avaliação em massa de imóveis usando geoestatística e krigagem bayesiana (um estudo de caso em Balneário Camboriú-SC).**Revista Eletrônica de Engenharia Civil ? REEC.**, v. 13, n. 1, 2017.
- [10] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **O Cidades** [online]. Disponível em: <<http://cod.ibge.gov.br/9EL>>. Acesso em 02/07/2017.
- [11] MARTÍNEZ, R. O.; ZAMPROGNO, B. Comparação de algumas técnicas de previsão em análises de séries temporais. **Revista Colombiana de Estadística** , v. 26, n. 2, p. 129-157, 2003.



- [12] MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 2 ed. Rio de Janeiro: LTC Editora, 2008. 463 p.
- [13] PEREIRA, J. C.; GARSON, S.; ARAÚJO, E. G. Construção de um modelo para o preço de venda de casas residenciais na cidade de Sorocaba-SP. **Gestão da Produção, Operações e Sistemas**, Bauru, v. 7, n. 4, p. 153-167, 2012.
- [14] **Prefeitura de Paranaíba**, 2017. Disponível em <<http://paranaiba.ms.gov.br/site/index.php>>. Acesso em 10 maio 2017.
- [15] R Core Team. **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria, 2017. <https://www.R-project.org/>.
- [16] SÁ, A. R. S. Avaliação Imobiliária: método comparativo de dados do mercado - tratamento científico. **Revista Especialize On-line IPOG** [on-line]. Edição 5. Goiânia: Instituto de Pós-graduação e Graduação - IPOG, 2013, Julho/2013. Disponível em: < <https://www.ipog.edu.br/revista-especialize-online/>>. Acesso em 03/07/2017.
- [17] SCHMIDT, A.M.; SANSÓ. B. **Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais**. Caxambu: ABE, 2006. 151 p.
- [18] STEINER, M. T. A.; CHAVES NETO, A.; BRAULIO, S. N.; ALVES, V. Métodos estatísticos multivariados aplicados à engenharia de avaliações. **Gestão da Produção, Operações e Sistemas**, Bauru, v. 15, n. 1, p.23-32, abr. 2008.
- [19] THOFEHRN, R. **Avaliação em massa de imóveis urbanos: para cálculo de IPTU e ITBI**. São Paulo: Pini, 2010.
- [20] TRIVELLONI, C. A. P.; HOCHHEIM, N. O valor de localização dos imóveis: determinação por métodos de análise espacial. In: IBAPE ? XXII UPAV / XIII COBREAP, 22., 2006, Fortaleza. **Anais...** . Fortaleza: Ibape, 2006. p. 1 - 40.

---

Submetido em 3 out. 2017

Aceito em 27 jul. 2018