

## **ANÁLISE DE MODELOS DE REGRESSÃO LINEAR MÚLTIPLA PARA ELABORAÇÃO DE PLANTA DE VALORES GENÉRICOS**

### *Analysis of Multiple Linear Regression Models to Produce Land Value Assessment Dataset*

**Renan Furlan de Oliveira**

**Universidade Estadual Paulista (UNESP) - Faculdade de Ciências e Tecnologia (FCT)**

Programa de Pós-Graduação em Ciências Cartográficas – PPGCC  
Rua Roberto Simonsen, 305, 19060-900 Presidente Prudente - SP - Brasil  
renanfurlanoliveira@hotmail.com

**Ana Paula Marques Ramos**

**UNOESTE – Universidade do Oeste Paulista - Faculdade de Ciências, Letras e Educação  
(FACLEPP)**

Departamento de Geografia  
Rua José Bongiovani, 700, 19 050-9205 - Cidade Universitária, Presidente Prudente - SP – Brasil  
anaramos@unoeste.br

**Carolina Scherrer Malaman**

**UNICAMP - Colégio técnico de Limeira**

Departamento de Construção Civil e Geomática  
Rua Pedro Zaccarias, 1300, Jardim Santa Luiza, Limeira- SP – Brasil  
cmalaman@gmail.com

**Amilton Amorim**

**Universidade Estadual Paulista (UNESP) - Faculdade de Ciências e Tecnologia (FCT)**

Departamento de Cartografia  
Rua Roberto Simonsen, 305, 19060-900 Presidente Prudente - SP - Brasil  
amorim@fct.unesp.br

#### **Resumo:**

O presente trabalho propõe a validação de modelos de Regressão Linear Múltipla (RLM) para a elaboração de Planta de Valores Genéricos (PVG). Dois modelos foram estimados, um para o conjunto amostral da variável de interesse não tendo distribuição normal e, o segundo modelo, com dados normais. Além da análise de variância (ANOVA) e dos testes de hipóteses recomendados pela ABNT (NBR 14.653-2, 2004), fez-se a validação dos modelos por meio da Raiz quadrada do Erro Médio Quadrático (REMQ) no Valor do Metro Linear de Testada Corrigida (VMLTC) de 11 pontos de verificação. A partir do modelo de menor REMQ, com base em simulação, avaliou-se a influência na arrecadação de impostos ao se utilizar a RLM proposta. Os modelos satisfazem aos testes de hipóteses, porém, ocorre uma significativa redução da REMQ na amostra com distribuição normal. No caso estudado, a análise pelo REMQ permitiu identificar o modelo que melhor representa a PVG, o que pode minimizar tanto as perdas financeiras quanto a injustiça fiscal no município.

**Palavras-chave:** Cadastro Técnico Multifinalitário, Planta de Valores Genéricos, Regressão Linear Múltipla, Análises Estatísticas.

#### **Abstract**

This work presents the validation of Multiple Linear Regression (MLR) models to the generation of a Land Value Assessment Dataset. Two models were estimated through a sample set, in which the first model was estimated using

**Anais do COBRAC 2016 - Florianópolis –SC – Brasil - UFSC – de 16 à 20 de outubro 2016**

raw data without normal probability distribution and the latter used normal data. The models were validated based on the Analysis of Variance (ANOVA) and on the hypothesis tests recommended by the ABNT (NBR 14.653-2, 2004). The models were also assessed through the RMSE (Root Mean Square Error) in 11 check points. The influence on tax revenue was assessed based on simulation using the model with the lowest RMSE. The models comply with the hypothesis tests. However, a significant reduction occurred in the RMSE when the normal distributed data was used. The RMSE analysis in the case studied allows identifying the model that best represents the PGV, which can minimize the financial losses or fiscal injustice in the city.

**Keywords:** Multipurpose Cadastre, Land Value Assessment Dataset, Multiple Linear Regression, Statistical Analysis.

## 1 INTRODUÇÃO

A partir da Lei 10357/01, que instituiu o Estatuto das Cidades, os municípios devem aprimorar o sistema de gestão de tributos e a sistemática de coleta, organização e disponibilidade de dados para um melhor planejamento (BARROS, 2011). Para uma gestão eficaz dos recursos de um município, tais como os recursos econômicos, o administrador público deve dispor de informações atualizadas que descrevam a região e, ser responsável pela tomada de decisões (FRANCISCO; IMAI, 2003). Um instrumento importante para o uso eficaz desses recursos é a Planta de Valores Genéricos (PVG), a qual compõe o Cadastro de Valores que representa uma das vertentes do Cadastro Técnico Multifinalitário (CTM) (DALAQUA *et al.*, 2010).

O CTM é definido como um conjunto de informações gráficas e descritivas provenientes de levantamentos cartográficos e socioeconômicos sobre uma porção da superfície terrestre, de maneira que estas informações possibilitem o conhecimento detalhado das propriedades imobiliárias, auxiliando na administração municipal à realização de uma gestão racional, legal e econômica dos recursos existentes (ERBA *et al.*, 2005). Dessa forma, uma PVG não é utilizada somente como instrumento de tributação e gestão de impostos, do tipo ITBI (Imposto de Transmissão de Bens Imóveis), ITU (Imposto Territorial Urbano), IPTU (Imposto Predial e Territorial Urbano), entre outros, mas, sobretudo, pode ser empregada como um recurso de promoção da justiça social e, respaldar a política de planejamento urbano, a qual visa garantir o desenvolvimento sustentável de uma cidade (AVERBECK, 2003; DALAQUA, 2007; SCHNEIDER, 1993).

Os valores dos objetos considerados em uma PVG consistem em terrenos e edificações (LIMA, 1999). No caso dos terrenos, pode-se representar o valor por metro quadrado, metro quadrado fixado por face de quadra ou adotando o Valor do Metro Linear de Testada Corrigida (VMLTC), o qual é determinado em função do valor do terreno e da testada (DALAQUA, 2007).

Segundo a Associação Brasileira de Normas Técnicas (ABNT NBR 14.653-2, 2004), o método utilizado para calcular o valor dos terrenos é dependente das informações que se dispõe e do nível de rigor que se deseja para a PVG. Entretanto, a avaliação dos imóveis, em geral, é realizada por dois métodos principais: método direto e indireto. Independentemente do método considerado, direto ou indireto, as formas de se calcular os valores do terreno são variadas, podendo-se citar a utilização de fatores de homogeneização, redes neurais, aplicação de geoestatística (MARQUES *et al.*, 2012) e regressão linear (DALAQUA *et al.*, 2010).

Neste trabalho, considerou-se o valor dos terrenos representados pelo VMLTC e estimados com base no método direto, especificamente o método comparativo de dados de mercado. A técnica estatística empregada sobre os dados consistiu em Regressão Linear Múltipla (RLM),

conforme a abordagem adotada por Dalaqua *et al.* (2010). A ABNT (NBR 14.653-1, 2001) caracteriza o método comparativo de dados de mercado como aquele que indica o valor do terreno a partir da realização de um tratamento técnico sobre os atributos dos elementos comparáveis (terrenos), os quais formam o conjunto amostral. Assim, realiza-se um tratamento sobre as características do objeto de interesse, tais como topografia, dimensões, localização e dados de comercialização. Estes últimos podem, por exemplo, serem obtidos na prefeitura ou até mesmo nas imobiliárias da cidade.

O objetivo deste trabalho consiste em analisar dois modelos de RLM a partir da análise de variância (ANOVA), dos testes de hipóteses recomendados pela ABNT (NBR 14.653-2, 2004) e por meio da Raiz quadrada do Erro Médio Quadrático (REMQ) sobre o VMLTC de 11 pontos, os quais foram selecionados como pontos de verificação dos modelos. Embora a ABNT (NBR 14.653-2, 2004) apresente os procedimentos para obter uma PVG por RLM, recomenda-se que a população de onde são retirados os dados a serem utilizados para ajustar o modelo tenha distribuição normal. Porém, muitas vezes, diante de uma variável dependente com distribuição muito próxima da normal, trabalha-se com dados como se fossem normais (DALAQUA *et al.*, 2010). Sendo assim, a principal contribuição deste trabalho é verificar, por meio da análise da REMQ no VMLTC dos pontos de verificação, qual a influência de se utilizar um modelo de regressão obtido por dados sem normalidade (amostra bruta) e dados com distribuição normal após uma transformação nos dados. Além disso, pretende-se utilizar o modelo de menor erro para realizar simulações sobre o impacto na arrecadação de impostos do município ao se produzir a PVG com base no modelo proposto.

## 2 ESTIMATIVA E VALIDAÇÃO DOS MODELOS DE REGRESSÃO

Para realizar a validação das estimativas dos valores da variável resposta e, posteriormente, a simulação sobre a arrecadação de impostos com o modelo estabelecido, foram realizadas as seguintes fases: delimitação da área de estudo, teste de normalidade do conjunto amostral, determinação dos modelos de regressão e validação das estimativas. A validação dos modelos ocorreu tanto pela análise de variância quanto pelos testes sugeridos pela ABNT (NBR 14.653-2, 2004). Além disso, determinou-se a Raiz quadrada do Erro Médio Quadrático (REMQ) no VMLTC considerando duas situações, inicialmente, sobre os pontos que participaram do ajuste dos modelos e, posteriormente, sobre os pontos selecionados para a verificação dos modelos matemáticos.

### 2.1 Área de estudo

A cidade de Álvares Machado localiza-se no oeste do Estado de São Paulo e conta com uma população de aproximadamente 23.513 habitantes e 8.493 imóveis urbanos (IBGE, 2010). Os dados aqui utilizados são resultados de um levantamento realizado por Dalaqua (2007) e referem-se aos valores (em reais) das transações imobiliárias de terrenos, ocorridas no ano de 2006 nesta cidade. As informações são provenientes de consultas realizadas junto às imobiliárias de Álvares Machado, bem como de entrevistas com os compradores e/ou vendedores de imóveis.

A variável resposta utilizada para a elaboração da PVG não consistiu simplesmente no valor de negociação dos terrenos, mas também em uma unidade de medida do valor, denominada de Valor do Metro Linear da Testada Corrigida (VMLTC), conforme a abordagem de Harper-Berrini (DALAQUA *et al.*, 2010). A estimativa do VMLTC é realizada em função do valor de comercialização do terreno e de informações descritivas que caracterizam o imóvel, as quais

foram obtidas por Dalaqua (2007) junto ao Banco de Dados da Prefeitura de Álvares Machado.

Utilizou-se um conjunto amostral de 77 observações, das quais 11 foram selecionadas para verificar as estimativas proporcionadas pelos modelos de regressão avaliados e, portanto, não participaram do ajuste dos modelos. A Figura 1 apresenta a área de estudo e a distribuição dos 66 pontos utilizados na determinação dos modelos, bem como os 11 pontos de verificação. Estes pontos de verificação foram selecionados de forma a apresentarem uma boa distribuição espacial na região de interesse.

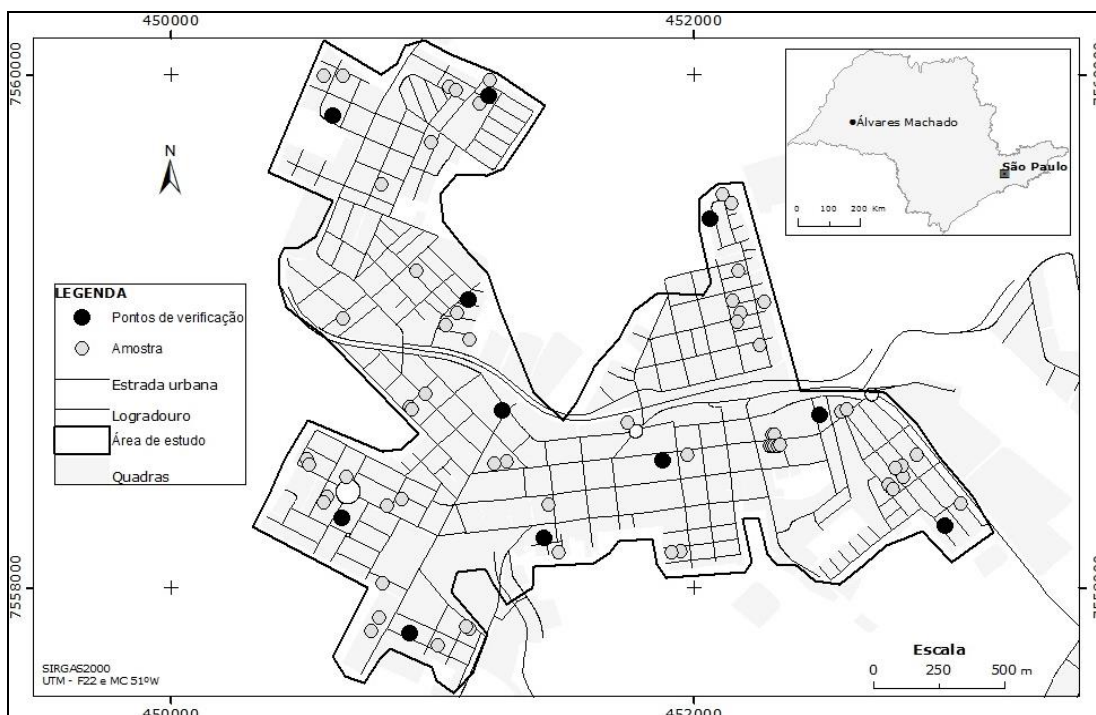


Figura 1 – Área de estudo e distribuição dos pontos amostrais.

## 2.2 Teste de normalidade

Para verificar se a variável resposta (VMLTC) apresenta distribuição de probabilidade normal, aplicou-se o teste de Anderson-Darling (AD) a um nível de significância  $\alpha$  (alpha) de 5% ( $p\text{-valor} < 0,005$ ) (Figura 2). A partir deste teste, constatou-se que os dados não possuem distribuição normal, o que justificou a aplicação de uma função de transformação na tentativa de se obter normalidade nos dados. Para isto, utilizou-se a função Box-Cox (Equação 1), considerando duas transformações de potências ( $\lambda$ ):  $\lambda = 0$  e  $\lambda = 0,333$ , em que  $x$  representa os valores da amostra (JOHNSON; WICHERN, 2002).

$$x^\lambda = \frac{x^\lambda - 1}{\lambda}, \lambda \neq 0 \quad \text{e} \quad x^\lambda = \ln x, \lambda = 0 \quad (1)$$

Ao analisar os valores amostrais versus os percentis esperados para dados normais, apresentados no gráfico Q-Q Plot (Figura 3), constatou-se que a transformação com potência  $\lambda = 0,333$  é o parâmetro que mais maximiza a função e, tem por consequência, a menor soma do quadrado residual. Portanto, melhor transforma a distribuição dos dados da amostra em uma

distribuição normal. A análise visual sobre o gráfico (Figura 3) permite aceitar a hipótese de normalidade, pois os pares de pontos encontram-se bem distribuídos ao longo da reta que indica a distribuição normal.

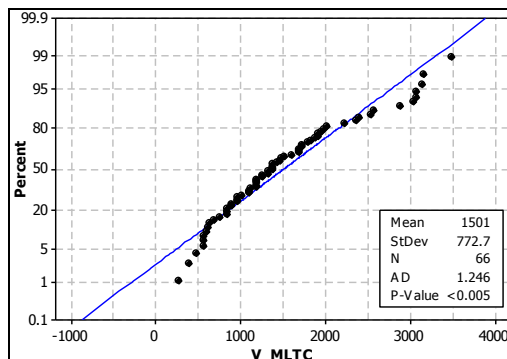


Figura 2 – Teste de normalidade realizado no *software* Minitab.

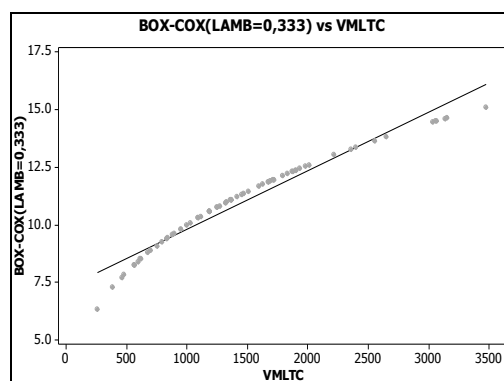


Figura 3 – Gráfico Q-Q Plot para  $\lambda = 0,333$ .

### 2.3 Determinação dos modelos de regressão

A análise de regressão sobre os dados imobiliários tem por objetivo determinar um modelo estatístico que descreva o relacionamento entre o valor do imóvel (variável resposta) e as variáveis que caracterizam a formação do valor do bem (variáveis independentes) (NADAL *et al.*, 2003). Para a estimativa dos dois modelos de regressão, considerou-se o mesmo número e tipo de variáveis independentes ou variáveis explicativas, utilizando-se um total de oito, as quais foram definidas por Dalaqua *et al.* (2010) como o melhor conjunto para descrever os terrenos na cidade de Álvares Machado (Quadro 1). Os modelos propostos para dados com e sem normalidade, determinados por meio do *software* Minitab 16®, estão apresentados no Quadro 2.

Quadro 1 – Variáveis independentes selecionadas para os modelos de regressão.

Variável	Descrição	Unidade de Medida ou Categoria
Testada (TESTADA)	Representa a largura do terreno	Metros
Distância (T_DIST)	Distância do terreno ao centro da cidade	
Padrão construtivo (T_PADRAO)	Padrão de construção da região à qual o terreno pertence	Alto (3,0), Médio-Alto (2,5), Médio (2,0), Médio-Baixo (1,5) e Baixo (1,0)

Topografia (T_TOPO)	Classificação do terreno	Plano (1,0), Aclive (0,9), declive (0,8), declive acentuado e irregular (0,7)
X (X_COORD)	Posição geográfica do terreno no sistema UTM.	Metros
Y (Y_COORD)		
Distância da via (DIST_VIA)	Representa a distância até a via principal de tráfego	
Zona Comercial (ZONA_COMER)	Identifica se o terreno pertence à zona de comércio	Sim (1), Não (0)

Quadro 2 – Modelos de regressão estimados para inferir o VMLTC.

Variável resposta	Modelo de Regressão Linear Múltipla
Com distribuição normal	$VMLTC = 918 + 0,0244*TESTADA - 0,00171*T\_DIST + 1,61*T\_PADRAO + 3,62*T\_TOPO - 0,000364*X\_COORD - 0,000099*Y\_COORD - 0,00118*DIST\_VIA + 2,35*ZONA\_COMER$
Sem distribuição normal	$VMLTC = 325369 + 16,4*TESTADA - 0,639*T\_DIST + 526*T\_PADRAO + 1498*T\_TOPO - 0,161*X\_COORD - 0,0335*Y\_COORD - 0,550*DIST\_VIA + 1165*ZONA\_COMER$

As estatísticas encontram-se na Tabela 1, a qual evidencia que ambos os modelos apresentam um coeficiente de determinação ( $R^2$ ) alto e similar, indicando que podem representar a variabilidade da variável dependente em mais de 88%. Contudo, esse comportamento não descaracteriza a necessidade de se validar as estimativas, para a qual se pretende aplicar a análise de variância, os testes propostos pela ABNT (NBR 14.653-2, 2004) e o cálculo da REMQ para o VMLTC. Cabe ressaltar que, os valores de erro padrão apresentados na Tabela 1 se referem à estimativa da variância do componente de erro dos modelos.

Tabela 1 – Estatísticas dos modelos de regressão linear múltipla.

Amostra	Coeficiente de determinação ( $R^2$ )	Coeficiente ajustado ( $R^2$ )	Erro padrão
Com normalidade (dados transformados)	88,1%	86,4%	0,727
Sem normalidade (dados originais)	89,5%	88%	267,426

## 2.4 Validação dos modelos de regressão linear múltipla

De acordo com Johnson e Wichern (2002), para que uma regressão possa ser utilizada na inferência de valores para uma variável dependente, é necessário que o modelo gerado satisfaça a alguns pressupostos estatísticos, a fim de se obter avaliações não-tendenciosas, eficientes e consistentes. Esta etapa é conhecida como validação do modelo. Neste sentido, inicialmente, realizou-se a análise de variância (ANOVA) dos modelos, com o intuito de verificar a hipótese de representatividade estatística de ambos.

Na sequência, aplicou-se os testes recomendados pela ABNT (NBR 14.653-2, 2004), a qual dispõe, em seu Anexo A, sobre os procedimentos para a utilização de modelos de regressão linear. Esta NBR ressalta a necessidade de analisar as estimativas dos erros (resíduos) obtidos pelo modelo de regressão de maneira a verificar, principalmente, a normalidade dos erros,

homocedasticidade, não-multicolinearidade entre as variáveis independentes, não-autocorrelação dos erros e inexistência de pontos atípicos (outliers). A validação dos modelos de regressão foi realizada em fases distintas e independentes com o uso do software Minitab 16® e está apresentada nas seções 2.4.1 a 2.4.5.

#### **2.4.1 Análise de variância (ANOVA)**

A análise de variância foi aplicada a um nível de significância de 5% e baseia-se no teste F (Fisher-Snedecor). O propósito do teste está em avaliar o quanto da variabilidade da variável de interesse (VMLTC) pode ser explicada pelas variáveis independentes. As hipóteses testadas foram  $H_0: \beta_0 = 0, \beta_1 = 0, \dots, \beta_p = 0$ , contra  $H_1$ : pelo menos um dos coeficientes do modelo é diferente de zero. Se  $H_0$  for rejeitada, conclui-se que há contribuição significativa de uma ou mais variáveis regressoras no estudo da variável resposta (JOHNSON; WICHERN, 2002).

Para o modelo de regressão estimado com base nos dados sem distribuição normal, obteve-se um F calculado de 60,71, enquanto que, para o modelo ajustado com dados normais esse valor de F foi de 52,82. Para o nível de significância considerado (5%) o valor tabelado de F, com oito e 57 graus de liberdade, foi de 3,15. Portanto, rejeita-se a hipótese  $H_0$  nos testes de ambos os modelos e, conclui-se, com 95% de confiança, que existem variáveis que auxiliam nessa análise, sendo, portanto, vantajosa a adoção dos modelos na inferência do VMLTC.

#### **2.4.2 Teste de multicolinearidade**

A identificação da correlação entre as variáveis independentes é importante, pois a ocorrência de uma forte dependência linear entre duas ou mais variáveis explicativas provoca degenerações no modelo e limita a sua utilização (ABNT NBR 14.653-2, 2004). Assim, para investigar a multicolinearidade, calcula-se uma matriz das correlações entre as variáveis independentes, que representa a dependência linear de primeira ordem entre essas variáveis.

Em geral, considera-se que a correlação entre as variáveis independentes deve ser inferior a 0,8. No caso de se encontrar valores superiores, a alternativa é a remoção de uma das variáveis altamente correlacionadas, para evitar dificuldades na predição da variável resposta (ABNT NBR 14.653-2, 2004). Neste contexto, o teste de multicolinearidade aplicado, considerando as oito variáveis explicativas, apontou que a maior correlação (0,63) ocorre entre as variáveis T\_DIST e Y\_COORD, o que descaracteriza a existência de um problema grave de multicolinearidade no conjunto de variáveis.

#### **2.4.3 Teste de normalidade dos erros**

A regressão linear baseia-se na hipótese de que os erros possuem uma distribuição de probabilidade normal e, em caso dessa hipótese ser refutada, tem-se que o modelo de regressão encontrado não é adequado (JOHNSON; WICHERN, 2002). Segundo a ABNT (NBR 14.653-2, 2004), a normalidade dos erros pode ser avaliada usando os testes de Anderson-Darling (AD) ou Ryan-Joiner (RJ), sendo estes indicados para um conjunto amostral pequeno como o conjunto amostral utilizado neste trabalho. Tanto o teste AD quanto o RJ, revelaram que os modelos ajustados a ambos os conjuntos de dados possuem erros com distribuição normal, a um nível de significância de 5%, visto que o p-valor foi superior a 0,05 em cada teste (Tabela 2).

Tabela 2 – Teste de normalidade dos erros por Anderson Darling e Ryan Joiner.

Dados	Anderson Darling	Ryan Joiner
Com normalidade	0,371 e p-valor = 0,431	0,992 e p-valor > 0,1
Sem normalidade	0,402 e p-valor = 0,349	0,991 e p-valor > 0,1

#### 2.4.4 Teste de homocedasticidade e correlação dos resíduos

Resíduos homocedásticos indicam a existência de uma variabilidade constante dos erros ao longo da reta de regressão (JOHNSON; WICHERN, 2002). Quanto ao teste de correlação dos resíduos, este possibilita avaliar se o resíduo de uma observação não está correlacionado com o resíduo de outra, e, se o modelo satisfizer este teste, tem-se que uma transação imobiliária não interfere na outra (DANTAS, 2005).

Para realizar tais testes, a ABNT (NBR 14.653-2, 2004) sugere a construção do gráfico entre os resíduos e os valores ajustados para a variável resposta. A homocedasticidade bem como a não correlação dos resíduos é caracterizada pela distribuição aleatória dos elementos residuais em torno do valor zero, comportamento este observado nos dois modelos de regressão propostos, como apresentam as Figura 4 e 5.

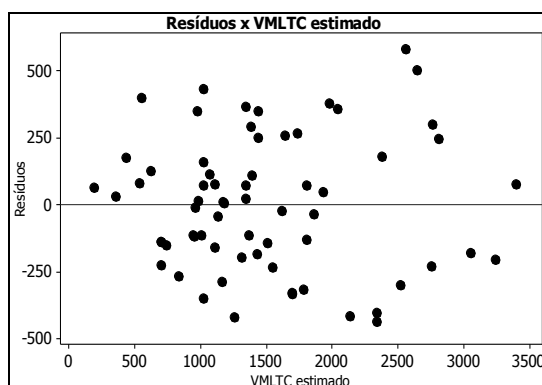


Figura 4 – Teste de homocedasticidade e correlação dos resíduos. Dados sem normalidade.

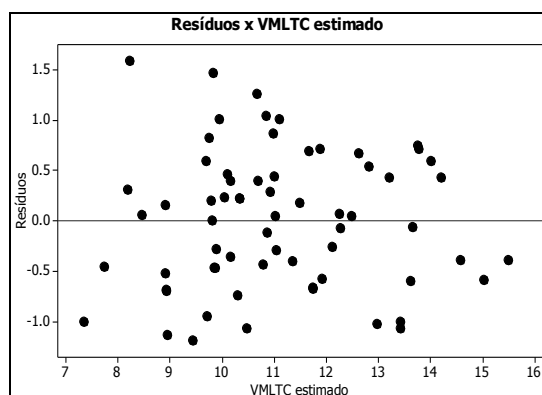


Figura 5 – Teste de homocedasticidade e correlação dos resíduos. Dados normais.

### 2.4.5 Teste de outliers

Dantas (2005) define um outlier como uma observação distante da média, resultado de um comportamento distinto comparado aos demais elementos amostrais. A existência desses pontos atípicos pode ser verificada pelo gráfico dos resíduos (ou resíduos padronizados) em relação aos valores ajustados (ABNT - NBR 14.653-2, 2004).

Ao se utilizar os resíduos padronizados, o resultado da distribuição dos pontos descaracteriza a presença de outliers se as observações estiverem compreendidas no intervalo de -2 a +2 desvios-padrão. Em caso de detecção de outlier, deve-se verificar os valores para identificar alguma inconsistência e remover tais pontos. As Figuras 6 e 7 apresentam dois possíveis outliers, todavia, decidiu-se mantê-los nos modelos de regressão, pois seus valores são condizentes com os terrenos que representam, após a realização de uma inspeção visual sobre a área que ocupam.

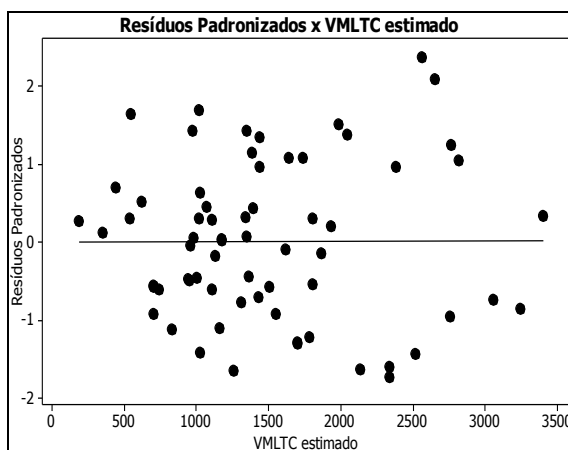


Figura 6 – Teste de *outliers*. Dados sem normalidade.

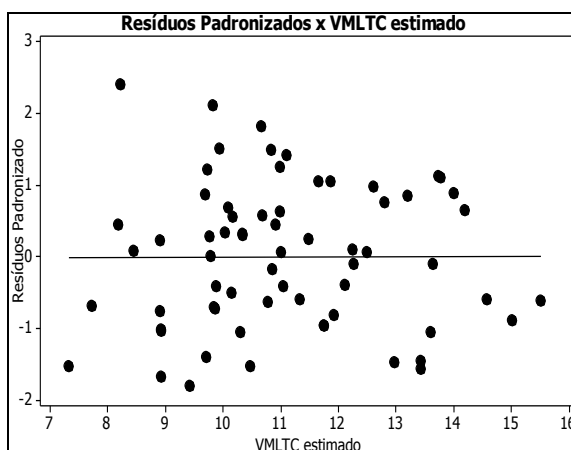


Figura 7 – Teste de *outliers*. Dados normais.

### 3 RESULTADOS E DISCUSSÃO

Além da análise de variância e dos testes de hipóteses recomendados pela ABNT (NBR 14.653-2, 2004), fez-se a validação dos modelos por meio da Raiz quadrada do Erro Médio Quadrático no VMLTC (Equação 2). Esse indicador considera as discrepâncias entre o valor amostrado em campo e o respectivo valor inferido pelo modelo matemático e é definido pela equação abaixo.

$$REM_{Q} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (VMLTC_{amostra\ i} - VMLTC_{estimado\ i})^2} \quad (2)$$

em que  $n$  é o número de observações.

Vale ressaltar que os valores estimados pelo modelo de regressão, obtido com os dados transformados pela Box-Cox por uma potência  $\lambda = 0,333$ , foram elevados ao cubo, uma vez que todas as observações (VMLTC) foram transformadas em mais ou menos raiz cúbica do valor original. Dessa forma, pode-se voltar à escala da variável original (dados brutos) e realizar a comparação da REMQ gerada pelos modelos de regressão. A primeira análise estimou a REMQ para todos os 66 pontos que participaram do ajuste dos modelos, cujos resultados são apresentados na Tabela 3.

Tabela 3 – Comparação da REMQ no VMLTC dos 66 elementos amostrais.

Amostra	REMQ no VMLTC em reais (R\$)
Com distribuição normal	259,80
Sem distribuição normal	250,43
Diferença da REMQ no VMLTC	9,37
Aumento da REMQ	$(9,37/250,43) * 100\% = 3,74\%$

Observa-se, na Tabela 3, que a discrepância da REMQ entre os modelos é baixa e inferior a 4%. Esse resultado aponta para uma grande semelhança nas estimativas do VMLTC, o que corrobora os resultados da análise de variância e justifica os valores similares dos coeficientes de determinação apresentados por ambos os modelos. Entretanto, a fim de avaliar a representatividade do modelo para os pontos que não foram utilizados na sua determinação, realizou-se uma segunda análise, a qual envolveu somente os 11 pontos de verificação (Tabelas 4 e 5).

Nesta análise, apresentada nas Tabelas 4 e 5, nota-se que a discrepância da REMQ entre os dois modelos é alta, chegando a aproximadamente R\$110,00, ou seja, existe uma capacidade distinta dos modelos em estimar valores para a variável resposta em locais aleatórios da área de estudo. Em termos dos percentuais (Tabela 6), isso implica que o erro na estimativa do VMLTC pela regressão oriunda dos dados com distribuição normal pode ser da ordem de 30% menor comparado à regressão gerada pelas amostras sem normalidade.

Tabela 4 - REMQ no VMLTC dos pontos de verificação para o modelo com distribuição normal.

Coordenadas		VMLTC (R\$)		
E (m)	N (m)	Estimado	Conhecido	Resíduos
451215	7559917	587,84	698,32	-110,48
450619	7559842	512,72	459,22	53,50
451136	7559122	1775,61	1272,73	502,89
450652	7558273	1661,67	1712,33	-50,66
450913	7557824	2011,49	1715,85	295,64
451427	7558193	2264,08	1871,68	392,40
451881	7558496	2520,18	2647,25	-127,07
452961	7558243	833,26	791,67	41,60
452061	7559440	900,93	751,88	149,06
452483	7558674	1143,00	1030,47	112,54
451268	7558692	2473,95	1631,21	842,74
REMQ				R\$357,05

Tabela 5 – REMQ no VMLTC dos pontos de verificação para o modelo sem distribuição normal.

Coordenadas		VMLTC (R\$)		
E (m)	N (m)	Estimado	Conhecido	Resíduos
451215	7559917	456,39	698,32	-241,93
450619	7559842	777,81	459,22	318,59
451136	7559122	1705,92	1272,73	433,187
450652	7558273	1266,22	1712,33	-446,107
450913	7557824	1388,55	1715,85	-327,299
451427	7558193	2485,95	1871,68	614,1695
451881	7558496	3032,10	2647,25	384,847
452961	7558243	1166,14	791,67	374,4685
452061	7559440	614,11	751,88	-137,769
452483	7558674	1412,08	1030,47	381,612
451268	7558692	2465,67	1631,21	834,464
REMQ				R\$466,94

Tabela 6 – Comparação entre os modelos de regressão nos pontos de verificação.

Modelo de regressão	REMQ (R\$)
Sem normalidade (dados originais)	466,94
Normal (dados transformados)	357,05
Diferença da REMQ no VMLTC	109,89
Aumento da REMQ	$(109,89/357,05)*100\% = 30,77\%$

Verifica-se que, embora os modelos apresentaram-se semelhantes nas análises de variância, assim como nos testes desenvolvidos nas seções 2.4.1 a 2.4.5, somente ao analisar as estimativas sobre os pontos de verificação é que se pode constatar, para o caso em estudo, que o

modelo ajustado a partir de dados normais proporciona uma melhor estimativa para a variável resposta, evidenciando a importância de se utilizar dados que seguem essa distribuição. Cabe ressaltar que se esperam valores superiores para os resíduos da REMQ nos pontos de verificação comparados aos resíduos da REMQ nos pontos usados no ajuste dos modelos, como foi constatado nas análises (Tabelas 4 e 5), pois esses elementos de verificação não são utilizados na determinação do modelo matemático.

Com a finalidade de realizar uma simulação sobre o impacto na arrecadação de impostos do tipo IPTU, por exemplo, ao utilizar o modelo de regressão de menor REMQ, realizou-se uma simulação do valor total de arrecadação com os novos valores. Todavia, como os dados utilizados referiram-se apenas às informações do terreno, desconsiderando, portanto, o valor das edificações, não se pode estimar diretamente a perda ou ganho de IPTU. Neste caso, adotou-se a taxa de 1% do valor venal do terreno sem considerar o valor venal das edificações. Assim, pode-se constatar o impacto medido no Imposto Territorial Urbano, que compõe o IPTU.

Com base nos valores obtidos para a totalidade dos dados, obteve-se uma variação de arrecadação de 23% entre os dois métodos. Segundo os dados da fundação SEADE, em 2007, data em que os dados foram coletados, a arrecadação de IPTU de Álvares Machado foi de R\$ 853.000,00. Considerando a variação encontrada nos dados, é importante perceber que, quanto maior for o cuidado tomado na definição do modelo e em seu ajuste, as chances de ocorrência de injustiça fiscal ou renúncia de receitas serão minimizadas. Percebe-se também pelos valores encontrados, a existência de um impacto significativo na arrecadação, que certamente tem consequências para a população.

#### **4 CONCLUSÕES E RECOMENDAÇÕES**

A pretensão desse trabalho foi caracterizar algum indicativo de qualidade quanto à inferência do VMLTC via modelos de regressão, estimados com dados que possuem ou não distribuição normal e apontar a influência dessa distribuição na arrecadação de impostos por meio de uma simulação. Embora os coeficientes de determinação tenham se mantido altos e semelhantes, além de ambos os modelos terem satisfeitos os testes da ABNT (NBR 14.653-2, 2004) e a análise de variância, constatou-se, neste conjunto de dados de estudo, a importância da validação da variável resposta pelo uso de pontos de verificação. Nessa análise, pode-se avaliar a diferença existente na capacidade dos modelos em estimar o VMLTC de terrenos distribuídos aleatoriamente na cidade e o erro (para mais ou para menos) que se comete nessas estimativas.

Assim, ainda que a amostra bruta tenha vindo de uma população com distribuição de probabilidade muito próxima à distribuição normal, no caso estudado, a análise da REMQ nos pontos de verificação aponta para a contribuição positiva ao se trabalhar com uma variável resposta normal ou transformada, pois, pode-se minimizar os erros nas estimativas da variável de interesse, o que reflete na confiabilidade da planta de valores gerada pela regressão e confirma as recomendações da ABNT (NBR 14.653-2, 2004). Isto é, a distribuição normal dos dados é importante para se ter uma boa estimativa do VMLTC.

Assim, este trabalho corrobora os estudos que utilizam a técnica de regressão linear múltipla para produzir planta de valores genéricos (DALAQUA, 2007; DALAQUA *et al.*, 2010; NADAL *et al.*, 2003). No entanto, aponta para a necessidade de validar as estimativas utilizando pontos que não participam do ajuste do modelo. Neste estudo, o modelo gerado a partir de dados normais foi capaz de estimar o VMLTC com uma REMQ 30% inferior comparado ao modelo ajustado para a variável resposta sem normalidade.

Com relação ao procedimento de simulação efetuado, constatou-se que o impacto sobre os tributos fiscais pode ser justo, sem onerar o contribuinte e nem proporcionar renúncia de receitas para a prefeitura. Ainda que, em muitos casos a arrecadação de IPTU, não seja a principal fonte de receita do município, ela é importante para outros programas de repasse do governo federal. Recomenda-se novos testes, em diferentes áreas e com mais elementos amostrais e de verificação. Sugere-se, ainda, elaborar uma PVG a partir da aplicação de outras técnicas de inferência, sobretudo, com característica espacial, com a finalidade de se determinar novas possibilidades para se gerar produtos ao CTM.

### **Referências Bibliográficas**

ABNT (Associação Brasileira de Normas Técnicas) (2004). **Avaliação de bens – Parte 2: imóveis urbanos** (NBR 14.653-2). Rio de Janeiro: ABNT.

ABNT (Associação Brasileira de Normas Técnicas) (2001). **Avaliação de bens – Parte 1: procedimentos gerais** (NBR 14.653-1). Rio de Janeiro: ABNT.

AVERBECK, C. E. **Os sistemas de cadastro e planta de valores no município: prejuízos da desatualização** (Dissertação). Universidade Federal de Santa Catarina – UFSC, Florianópolis, SC. 2003.

BARROS, G. H. S. **Sistemas cadastrais 3D: viabilidades para o sistema cadastral brasileiro** (Tese). Faculdade de Ciências e Tecnologia - FCT/UNESP, Presidente Prudente, SP. 2011.

DALAQUA, R. R. **Aplicação de métodos combinados de avaliação imobiliária na elaboração de valores genéricos** (Dissertação). Faculdade de Ciências e Tecnologia - FCT/UNESP, Presidente Prudente, SP. 2007.

DALAQUA, R. R.; AMORIM, A.; FLORES, E. Utilização de métodos combinados de avaliação imobiliária para a elaboração da planta de valores genéricos. **Boletim de Ciências Geodésicas**. v. 16, no 2, p.232-252, 2010.

DANTAS, R. A. **Engenharia de avaliações: uma introdução à metodologia científica**. São Paulo: Pini, 2005. Ed 2.

ERBA, D. A.; OLIVEIRA, F. L.; JUNIOR, P. N. L. **Cadastro Multifinalitário como instrumento de política fiscal e urbana**. Rio de Janeiro. 2005.

LIMA, O. P. **Proposta metodológica para o uso do Cadastro Técnico Multifinalitário na Avaliação de Impactos Ambientais** (Dissertação). Universidade Federal de Santa Catarina - UFSC. Florianópolis, SC. 1999.

FRANCISCO, H. R.; IMAI, N. N. Temporalidade em SIG: Aplicação em cadastro de imóveis. **Boletim de Ciências Geodésicas**. v. 9, n 1, p. 89-104, 2003.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. 2010. Disponível em



12º Congresso de Cadastro Técnico Multifinalitário e Gestão Territorial  
10º Encontro de Cadastro Técnico Multifinalitário para os países do Mercosul  
7º Encontro de Cadastro Técnico Multifinalitário para os países da América Latina  
ISBN 1980 - 4520

<<http://www.ibge.gov.br>>. Acesso em junho de 2016.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice-Hall, 2002. 5ª Ed.

MARQUES, A. P. S.; MARCATO JUNIOR, J.; AMORIM, A.; FLORES, E. F. Aplicação do Interpolador Krigagem Ordinária para a Elaboração de Planta de Valores Genéricos. **Revista Brasileira de Cartografia**, Nº 64/2, p. 175-186. 2012.

NADAL, C. A.; JULIANO, K. A.; RATTON, E. Testes estatísticos utilizados para a validação de regressões múltiplas aplicadas na avaliação de imóveis urbanos. **Boletim de Ciências Geodésicas**. v. 9, n 2, p. 243-262, 2003.

SCHNEIDER, V. P. **Análise do Sistema Tributário Quanto a Propriedade Imobiliária a Nível Municipal, Utilizando Dados do Cadastro Técnico Multifinalitário** (Dissertação). Universidade Federal de Santa Catarina – UFSC, Florianópolis, SC. 1993.