

EMPREGO DA ESTATÍSTICA MULTIVARIADA COMO PROPOSTA PARA O CÁLCULO DO VALOR VENAL E TRIBUTAÇÃO IMOBILIÁRIA

MULTIVARIATE STATICS EMPLOYED AS PROPOSAL FOR CALCULATING THE MARKET VALUE AND PROPERTY TAXATION

Jonilson Heil* E-mail: johniheil@yahoo.com.br
Neida Maria Patias Volpi* E-mail: neida@ufpr.br
*Universidade Federal do Paraná (UFPR), Curitiba - PR

Resumo: É conhecida a busca pelas municipalidades brasileiras em aumentar sua arrecadação própria e reduzir a dependência dos repasses financeiros estaduais e federais, otimizando suas receitas tributárias. Sabe-se também que os municípios pretendem realizar tal tarefa com idoneidade, clareza e facilidade de prestação de contas aos órgãos reguladores, bem como às suas respectivas populações. Neste trabalho procedeu-se um estudo sobre a metodologia empregada em um município da região centro-sul do estado paranaense para cálculos dos valores venais de imóveis, e conseqüente tributação do IPTU e ITBI sob estes bens. Ainda, com base nos dados cadastrais imobiliários cedidos por tal município, foi realizada, por meio de técnicas estatísticas multivariadas, uma análise das características que mais influem nas valorizações pecuniárias dos imóveis, e aplicando análise de regressão linear múltipla são propostos modelos de cálculo para estimação dos valores venais, possibilitando prognosticar cálculos tributários por seu intermédio. Finalmente são apresentadas comparações entre os resultados advindos da metodologia usada pelo município com os obtidos pelos modelos desenvolvidos, propostos para utilização em geral.

Palavras-chave: Tributação imobiliária municipal. Avaliação de imóveis. Estatística multivariada. Regressão linear múltipla.

Abstract: It is well known that the Brazilian municipalities aim to increase their own revenues and reduce dependence on state and federal financial transfers, optimizing their tax revenues. It is also known that the municipalities intend to carry out that mission with integrity, clarity and to present easily the accountability to regulators, as well as to their respective populations. In this paper carried out a study on the methodology employed in a town in central-southern state of Paraná to calculate the venal values and property tax (IPTU) and the consequent taxation of IPTU and ITBI in these goods. Based on municipality registration data was developed, by means of multivariate statistical techniques, an analysis of the characteristics that most influence the monetary valuations of the property, and applying multiple linear regression analysis are proposed models to estimate values of the venal values of properties, allowing tax calculations predict through it. Finally, comparisons are presented between the results from the methodology used by the municipality with those obtained by the models developed, proposed for use in general.

Keywords: Municipal property taxation. Property valuation. Multivariate statistics. Multiple linear regression.

1 INTRODUÇÃO

O Imposto Predial Territorial Urbano – IPTU – é um tributo de competência municipal, que tem como fato gerador a propriedade, o domínio útil ou a posse de bens imóveis localizados em área urbana. A lei que disciplina a matéria no Brasil é o Código Tributário Nacional (CTN), definindo que a base de cálculo do tributo é o valor venal do imóvel, e cada prefeitura o faz por meio de legislação própria que trata do conjunto de taxas e impostos sob sua competência.

O estudo do cálculo do IPTU justifica-se pela comum falta de atualização e uso de meios obsoletos e em alguns casos sem embasamento científico pelos municípios brasileiros. A ausência de atualizações neste sentido traz redução de receitas às prefeituras, bem como pode provocar cobranças discrepantes do tributo. As prefeituras buscam aliar aumento da arrecadação advinda de tributos próprios com transparência nas cobranças e satisfação de seus contribuintes quanto à clareza que este processo requer. Quando o foco desta discussão são o IPTU e o ITBI (Imposto Sob a Transmissão de Bens Imóveis), ambos tributos municipais calculados com base no valor venal dos imóveis, fica aparente a necessidade de se estabelecer métodos idôneos e ao mesmo tempo que possibilitem a tributação em conformidade com as tendências atuais de mercado, com cobranças de impostos justas e embasadas cientificamente. A cobrança do ITBI é oriunda da transferência de proprietário de bens imóveis por ato oneroso.

Outro fator que impulsiona essa pesquisa é a forma subjetiva de avaliação de imóveis desempenhada pelo mercado imobiliário, que em alguns casos pode exercer influência no momento de se fazer juízo de valor ao tributo, podendo trazer ao contribuinte a idéia de supervalorização ou até mesmo pensamento de defasagem abaixo do valor de mercado “justo” do imóvel. A existência de casos análogos ao citado podem deixar o município vulnerável à comparações naturais na cobrança dos impostos pela própria população.

2 TRIBUTAÇÃO IMOBILIÁRIA

Para a definição do valor a ser cobrado de IPTU e ITBI do contribuinte, as municipalidades aplicam uma alíquota sobre o valor de venda dos respectivos

imóveis. No entanto o CTN não especifica o modo pelo qual devem ser definidos os valores venais, deixando esta tarefa a cargo dos municípios.

Diante desta autonomia, cada município estipula os meios de obtenção do valor venal através de leis municipais próprias, normalmente denominadas Código Tributário Municipal (CTM). As regras de cobrança são definidas por tais leis que devem explicitar a composição dos cadastros imobiliários das prefeituras e as técnicas usufruídas na obtenção dos valores venais dos imóveis. Em cima disso são definidas as alíquotas incidentes e geradoras dos impostos em tela. Comumente é usado 1% do valor venal do imóvel para definição do IPTU para imóveis edificados, e 2% para não edificados. No ITBI geralmente prevalece a alíquota de 2% do valor venal do imóvel.

Matos & Portella (2005) menciona que “há afirmação doutrinária no sentido de que na prática, são atribuídos valores irrisórios à propriedade imobiliária – 10, 20 ou 30% do valor do mercado – e, em contrapartida, são fixadas alíquotas exorbitantes”. Se isso for praticado não somente o princípio de equidade tributária pode ser comprometido, como também impactar em uma menor arrecadação municipal, principalmente face a não observância do princípio da capacidade contributiva dos proprietários.

O cálculo do valor venal feito pelos municípios ocorre com base nas características e atributos que o bem recebe nos seus cadastros, após fiscalização *in loco*. Estas informações em geral são compostas pelas variáveis apresentadas nas tabelas do apêndice I.

Além dos CTM's, as municipalidades realizam um planejamento geral sobre as políticas de ocupação do município. Neste planejamento, chamado de Plano Diretor e previsto pela Constituição Federal em seu art. 182, normalmente deve haver algum tipo de mapeamento urbano que contemple, por meio de estudos detalhados das regiões envolvidas, as valorizações das propriedades de acordo com suas respectivas localizações (PERNAMBUCO, s.d.). Tais estudos detalhados devem ser elaborados no final de cada exercício fiscal, para que já no início do próximo ano estejam divulgados e em vigor.

Fica mais evidente neste ponto que qualquer cobrança de impostos de um imóvel depende do seu valor estimado e da maneira que o município o faz.

3 TRABALHOS CORRELATOS E DESCRIÇÃO DO PROBLEMA

O interesse crescente nos meios de avaliação de imóveis faz surgir diversos estudos na área. A necessidade eminente por justas maneiras de obtenção da valorização dos bens, aliada às evoluções tecnológicas, disseminação do conhecimento e a busca constante por aperfeiçoamento de técnicas, são fatores que despertam atenção de estudiosos destas áreas. O trabalho realizado por Alves (2005) propõe meios de avaliação com base na estatística multivariada. O artigo de Steiner et al. (2008) foca o assunto com o uso de meios de classificação aliada a regressão estatística na tarefa das avaliações dos bens imobiliários.

Uma justificativa para a este tipo de estudo é reportada no trabalho de González (2002), ao citar a importância econômica social do mercado imobiliário em uma região, mencionando que estimações dos valores venais dos bens são úteis em diversas ocasiões, como na liberação de financiamentos, estudo de viabilidade de novas construções, tributação, demandas judiciais e inventários. González (2002) propõe o cálculo das estimações na valorização de imóveis por intermédio de algumas técnicas estatísticas e numéricas, visando reduzir a subjetividade presente nos procedimentos comumente usados.

O trabalho de Skidmore et al. (2010) analisa as modificações ocorridas na forma de avaliação dos bens no Estado de *Michigan*, nos Estados Unidos, como consequência de uma nova lei sobre a matéria, inserida no ano de 1995, e, com base em informações coletadas junto à população no ano de 2008, verifica a redistribuição dos valores tributados de acordo com grupos demográficos e econômicos. Estimções dos valores de imóveis também são abordadas no trabalho de Nguyen & Cripps (2001). Trata-se de uma pesquisa de avaliação realizada com mais de 3000 observações coletadas no Estado norte-americano do *Tennessee*, efetuada por Regressão Linear Múltipla e Redes Neurais. Os autores comparam os resultados das técnicas usadas, enaltecendo a boa performance de todas as metodologias aplicadas.

O intuito preponderante nas pesquisas no âmbito da engenharia de avaliações, a exemplo do que ocorre na presente pesquisa, é o aprimoramento das técnicas e sugestões de seu emprego na melhoria das estimções dos valores imobiliários. Assim, embora a essência desta pesquisa esteja vinculada para fins

tributários, os estudos antes realizados fomentam e norteiam os objetivos centrais também no presente trabalho, onde é focado o desenvolvimento de metodologias para se determinar o valor venal de imóveis com base em informações extraídas de grandes bancos de dados, como na maior parte dos cadastros municipais, que possuem diversas variáveis influentes nas valorizações.

4 METODOLOGIA APLICADA

Para se alcançar o objetivo almejado foram utilizadas técnicas estatísticas multivariadas. A estatística multivariada é uma área de conhecimento da estatística que, por meio de suas técnicas, proporciona investigar cientificamente fenômenos que englobam diversas variáveis influentes, possibilitando estudar as relações entre as variáveis de forma integrada e evoluída, em geral fornecendo resultados consistentes (FERREIRA, 2008; JOHNSON & WICHERN, 1998; BAKKE *et al.*, 2008). As usadas nesta pesquisa são Análise de Regressão Linear Múltipla, Análise de Componentes Principais (ACP) e Análise de Agrupamentos (*Cluster Analysis*).

A análise de regressão linear trata de prognosticar o valor de uma variável dependente (resposta) por intermédio de sua relação com um conjunto de outras variáveis independentes (preditoras), bem como estudar o grau de influência que determinadas variáveis exercem umas nas outras (JOHNSON & WICHERN, 1998; BAKKE *et al.*, 2008; MARQUES & MARQUES). O modelo de regressão linear é dado por:

$$\underline{Y} = X \cdot \underline{\beta} + \underline{\varepsilon} \quad (1)$$

com

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \quad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix}_{n \times p} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}_{p \times 1} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

onde que X é a matriz do modelo, \underline{Y} o vetor das respostas, $\underline{\beta}$ o vetor dos parâmetros (coeficientes) e $\underline{\varepsilon}$ o vetor dos resíduos em uma população com p variáveis e n observações.

A análise de componentes principais (ACP) é uma técnica estatística multivariada que busca, dentro de um conjunto de variáveis que explicam um

fenômeno, aquelas que exercem maior influência nos resultados. Em uma população p -variada, se faz necessária a utilização de todas as p variáveis para uma reprodução íntegra da variabilidade existente em sua estrutura de covariância. No entanto comumente a maior parte desta variabilidade pode ser explicada por um número menor m de componentes principais (FERREIRA, 2008; LOPES & SAMOHYL, 2003; VICINI & SOUZA, 2005, MARQUES & MARQUES, 2005).

A idéia central da ACP está em evidenciar, dentro de um grupo de variáveis correlacionadas, quais as variáveis latentes que, combinadas linearmente a outras, podem explicar a maior parte da variação do modelo estudado, transformadas em novas variáveis através de combinações lineares, sendo estas novas não-correlacionadas. Após identificadas as novas variáveis latentes são ordenadas em função de suas variâncias, sendo definida como a primeira aquela que possuir combinação linear de maior variância, a segunda aquela com a segunda combinação de maior variância, e assim sucessivamente até a última variável abordada no modelo. Em seguida deve ser usado algum critério de seleção das variáveis mais importantes de modo a formar um conjunto pequeno de novas variáveis, de maneira que tal conjunto explique a maior proporção de variância do fenômeno analisado.

A análise de agrupamento é uma técnica estatística de classificação. Visa separar os itens focados pela análise conforme suas similaridades, dividindo a amostra trabalhada em grupos, alocando os itens considerados com pouco ou nenhum grau de semelhança em grupos distintos (JOHNSON & WICHERN, 1998; ROMESBURG, 2004; VICINI & SOUZA, 2005).

Por meio destas técnicas sucintamente descritas, a metodologia aqui proposta dá-se de acordo com os seguintes passos, que podem ser averiguados pelo fluxograma da FIGURA 1:

1º) Seleção e segregação de observações: visto que os terrenos baldios são identificados por menos e diferentes variáveis que os edificadas na determinação de sua valorização de venda, devem ser desempenhadas duas análises distintas para definição dos valores venais, uma para terrenos baldios e outra para lotes com edificações. Tal procedimento inclusive é previsto pela própria NBR 14653-2, norma brasileira criada pela ABNT que disciplina, em sua parte 2, as regras de avaliação de bens imóveis urbanos no país. Nesta etapa, além de separar as observações, já

devem ser excluídas do banco de dados as que estiverem incompletas (imóveis com variáveis/informações faltantes no cadastro);

2°) Definição das variáveis relevantes: como os bancos cadastrais imobiliários municipais geralmente contam com muitas variáveis, e sendo que algumas delas são meramente figurativas, nesta etapa deve-se definir quais serão as informações/características dos imóveis que deverão ser, inicialmente, incluídas nas análises. As nomenclaturas usadas pelos municípios também variam, no entanto a definição de quais variáveis serão utilizadas pode partir das que o próprio município estudado aplica, aliado às usadas nas práticas do mercado imobiliário regional;

3°) Aplicação da Regressão Linear Múltipla: de posse do banco de dados inicial a ser utilizado, são aplicadas diretamente duas análises de regressão linear múltiplas, para imóveis edificados e não edificados, para as observações e variáveis antes selecionadas, com variável resposta no valor venal do bem. Estas análises de regressão devem ser examinadas, e os resultados fornecidos subsidiarão decisões com respeito à suas aplicabilidades e possibilidades de aperfeiçoamento, se houver;

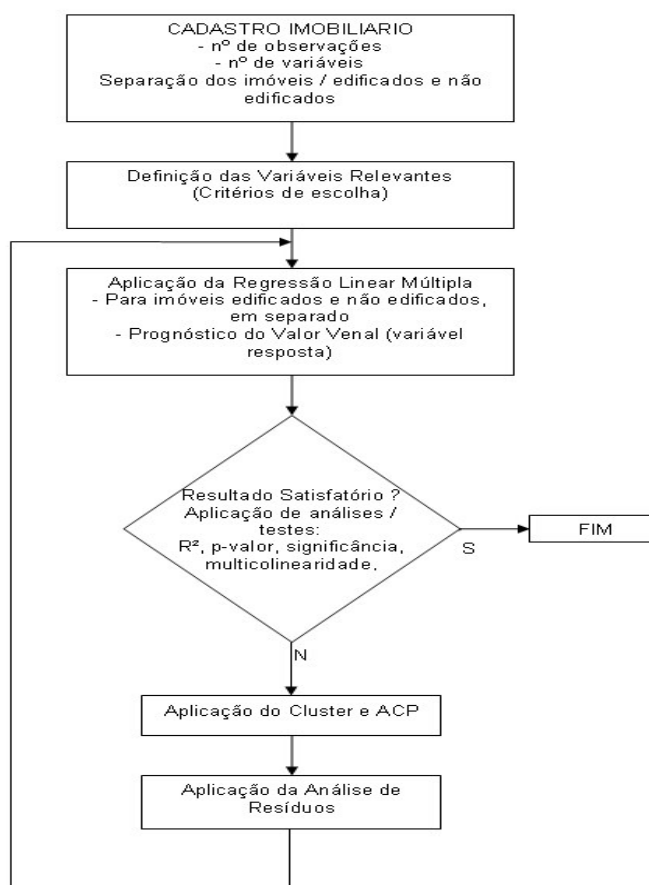
4°) Verificação dos resultados e consistência das análises: nesta etapa, com base nos resultados providos pelas regressões, deve ser definida a adoção ou não dos modelos calculados. Com subsídio nos testes desempenhados (R^2 , correlações, p-valor) é possível conhecer o poder de explicação dos modelos, possibilidades de aperfeiçoamento ou até mesmo descartar a sua utilidade. Contudo competem ao analista tais decisões. Caso entenda que os modelos poderão ser usados, porém exista a possibilidade de aprimoramentos, a análise prossegue, e o trâmite flui para o próximo passo. Do contrário, caso julgue os resultados consistentes e satisfatórios, os modelos poderão ser adotados para estimação dos valores venais dos imóveis e, sendo aplicadas as alíquotas pertinentes, calculados os tributos incidentes, encerrando as análises;

5°) Aplicação da ACP e Cluster: com intuito de aprimorar os modelos anteriores, são aplicadas duas análises, ACP e Agrupamentos, a fim de se conhecer melhor o comportamento das variáveis envolvidas. Tais análises podem apontar variáveis pouco contributivas, fomentando possibilidades de redução dos modelos, bem como destacar alguma semelhança de interesse nos comportamentos das variáveis, sugerindo estudá-las em conjuntos/grupos;

6°) Análise dos resíduos: erros muito elevados podem influir contundentemente os resultados de um modelo de regressão. Visando evitar esta influência negativa, um olhar de minúcia sobre os resíduos é desempenhada nesta etapa, tratando por *outliers* as observações detentoras dos maiores erros calculados, e eliminando-as da amostra trabalhada. Vale salientar que, em alguns casos, padrões e similaridades são identificados entre as observações dadas por *outliers*, o que poderá exigir alguma modificação no modelo ou interferência na forma como são tratadas estas observações.

7°) Eliminação de variáveis e observações: aqui são executadas as exclusões das variáveis vistas como supérfluas pelas análises anteriores, bem como das observações definidas como *outliers*, para que novos modelos de Regressão Linear Múltipla sejam criados a partir da amostra remanescente, retornando o fluxo da análise ao terceiro passo.

Figura 1 – Fluxograma das análises desempenhadas



5 APLICAÇÃO E RESULTADOS

O modelo proposto foi aplicado conforme o fluxograma da FIGURA 1, utilizando dados reais cedidos pela Prefeitura de um município localizado no centro-sul do Estado do Paraná. Inicialmente foram consideradas todas as observações do Cadastro Imobiliário Municipal, o que representou uma amostra composta por 10.068 imóveis. Foi ainda apurado a existência de 32 atributos cadastrados para cada imóvel que serviram, a princípio, como matéria-prima para variáveis do modelo. Com esta amostra foi construída uma matriz global de informações, contemplando todos os imóveis por linhas e seus respectivos atributos por colunas. Esta matriz com ordem 10.068X32 foi batizada de CARACTIMOV. A separação dos imóveis edificadas dos terrenos baldios dividiu a CARACTIMOV gerando duas outras matrizes, a CARACTIMOV BALDIOS, englobando apenas lotes vagos, de ordem 3.255X13, e a CARACTIMOV EDIFICADOS, comportando imóveis construídos ou em obras, de ordem 6.813X32.

Para seleção das variáveis a serem usadas nos modelos o critério foi manter os atributos mais considerados pelo mercado imobiliário na definição dos valores de venda dos bens. Ainda procurou-se preservar as variáveis usadas na metodologia da própria prefeitura, à luz do CTM daquele órgão. Estão distinguidas detalhadamente no apêndice I, e no modelo ficaram representadas da seguinte maneira:

- Variáveis dependentes: Y_T → valor venal apenas do terreno; Y_P → valor venal predial, apenas da edificação/construção; Y_I → valor venal total do imóvel ($Y_T + Y_P$).

- Variáveis independentes: X_1 → área do lote/terreno; X_2 → Índice de localização; X_3 → topografia; X_4 → pedologia; X_5 → situação; X_6 → fração ideal; X_7 → testada; X_8 → pavimentação; X_9 → área construída; X_{10} → acabamento externo; X_{11} → acabamento interno; X_{12} → cobertura; X_{13} → piso; X_{14} → esquadrias; X_{15} → estrutura; X_{16} → forro; X_{17} → instalação elétrica; X_{18} → instalação sanitária; X_{19} → revestimento externo; X_{20} → revestimento interno; X_{21} → número de pavimentos; X_{22} → localização horizontal; X_{23} → localização vertical; X_{24} → conservação.

O *software* estatístico utilizado para as análises desejadas foi o *Statgraphics Centurion*.

5.1 Análise para imóveis não edificadas (terrenos baldios)

Aqui a intenção é obter uma função linear que explique o relacionamento entre a variável dependente Y_T (valor venal do terreno - VVT), com as independentes referentes aos lotes baldios, X_1, X_2, \dots, X_8 .

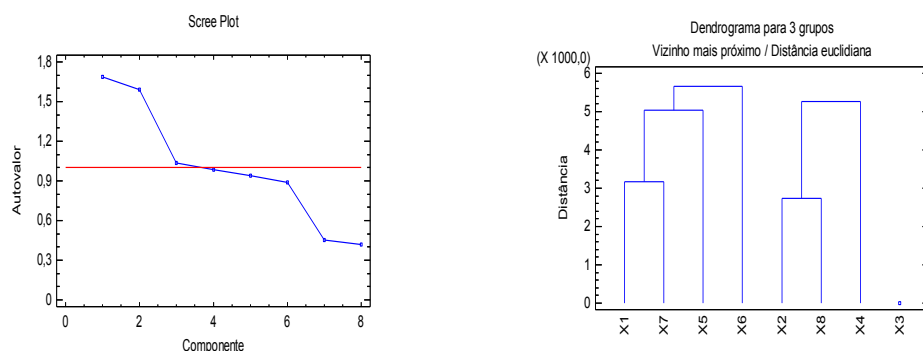
Foi realizada uma análise de regressão linear inicial, contemplando todos os dados da CARACTIMOV BALDIOS, e embora o modelo resultante tenha fornecido um coeficiente de correlação múltipla aceitável, com grau de explicação da variabilidade total em torno de 80 %, também gerou diversos *outliers*, e ainda correlações elevadas entre algumas variáveis. Isso representou possibilidade de aperfeiçoamento, e fomentou as análises de componentes principais e agrupamentos das variáveis, bem como tratamento dos *outliers*. Com todas as análises desempenhadas, uma nova Regressão Linear Múltipla foi executada e está apresentada em (2), além dos demais resultados que estão na seqüência.

$$Y_T = -591,415 + 0,0259452 \cdot X_1 + 444,629 \cdot X_2 + 128,262 \cdot X_3 + 82,5886 \cdot X_4 + 324,083 \cdot X_5 + 3,48082 \cdot X_7 \quad (2)$$

Quadro 1 – Resultados da ACP para terrenos

Principal Components Analysis			
Data variables: X1 (ÁREA DO LOTE), X2 (IND. LOC.), X3 (TOPOGRAFIA), X4 (PEDOLOGIA), X5 (SITUAÇÃO), X6 (FRAÇÃO IDEAL), X7 (TESTADA), X8 (PAVIMENTAÇÃO)			
Data input: observations extracted: 3	Standardized: yes	Number of components	
Principal Components Analysis			
Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	1,68486	21,061	21,061
2	1,59268	19,908	40,969
3	1,03575	12,947	53,916
4	0,986134	12,327	66,243
5	0,940142	11,752	77,995
6	0,888806	11,110	89,105
7	0,45299	5,662	94,767
8	0,41864	5,233	100,000

Figura 2 – Scree Plot da ACP e dendrograma (*Cluster*) para as variáveis dos terrenos em três grupos



s

Os resultados da ACP estão apresentados no QUADRO 1 e na FIGURA 2. Esta análise apontou que todas as variáveis exercem influência significativa nas respostas. Para uma explicação superior a 80% da variabilidade existente seria necessária a utilização de pelo menos 6 componentes, de um conjunto total de 8 variáveis.

Por sua vez, a análise de *cluster*, procedida pelo método do vizinho mais próximo, veio a demonstrar, corroborando com os resultados da ACP, quais variáveis possuem mais similaridade entre si, e em caso de separação por grupos, quais seriam mantidas juntas para fornecimento de resultados, conforme demonstrado no dendrograma da FIGURA 2.

Os resultados fornecidos confirmam ainda a utilidade do modelo (2) quanto ao teste de significância, pela estatística *F*, admitindo-se a hipótese de existência de relação linear entre a variável resposta (valor venal do terreno) com as variáveis predictoras. Percebeu-se também, pela matriz de correlação, não haver possibilidade de multicolinearidade no modelo.

Quadro 2 – Resultados da regressão linear múltipla aprimorada para terrenos**Multiple Regression - YT**

Dependent variable: YT (VVT EM UFM's)

Independent variables: X1 (ÁREA DO LOTE), X2 (ÍNDICE DE LOC.), X3 (TOPOGRAFIA), X4 (PEDOLOGIA), X5 (SITUAÇÃO), X7 (TESTADA)

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1,5281E8	6	2,54684E7	2592,79	0,0000
Residual	3,03032E7	3085	9822,75		
Total (Corr.)	1,83113E8	3091			

R-squared = 83,4511 percent

Standard Error of Est. = 99,1098

Mean absolute error = 52,1768

Durbin-Watson statistic = 1,45661 (P=0,0000)

Correlation matrix for coefficient estimates

	CONSTANT	X1	X2	X3	X4	X5	X7
CONSTANT	1,0000	-0,0271	0,0265	-0,3163	-0,5427	-0,7735	0,0674
X1	-0,0271	1,0000	0,0513	0,0217	0,0079	0,0405	-0,4320
X2	0,0265	0,0513	1,0000	0,0553	-0,1769	0,0008	0,0299
X3	-0,3163	0,0217	0,0553	1,0000	0,0344	-0,0214	0,0757
X4	-0,5427	0,0079	-0,1769	0,0344	1,0000	0,0031	-0,0176
X5	-0,7735	0,0405	0,0008	-0,0214	0,0031	1,0000	-0,1989
X7	0,0674	-0,4320	0,0299	0,0757	-0,0176	-0,1989	1,0000

O coeficiente de correlação múltipla calculado para o modelo em (2) apresentou um índice melhor do que o modelo inicial, demonstrando que agora cerca de 83,45% da variabilidade total existente é explicada.

Na análise dos resíduos pode-se reparar a diminuição dos erros, obviamente em virtude do tratamento e exclusão dos *outliers*., o que culminou na exclusão de 131 imóveis reduzindo a amostra para 3.091 observações. Ainda, as influências das observações possuidoras de grandes erros nos resultados foram eliminadas do modelo. Pelo gráfico de resíduos *versus* VVT observado nenhum indício de padronização é percebido na disposição dos pontos, possibilitando descartar a presença de correlação serial no modelo. É perceptível, pelo aspecto do mesmo gráfico, que os resíduos não apresentam nenhuma evidência de comportamento tendencioso (crescimento, decrescimento ou oscilação), significando que qualquer indício de heterocedasticidade também é descartado na regressão obtida.

Cabe ressaltar que o modelo em (2) ainda proporcionou redução de variáveis trabalhadas, descartando as variáveis X_6 (fração ideal) e X_8 (pavimentação) da análise. Este descarte pode também encontrar justificativa pelos resultados antes alcançados pela ACP e *Cluster*, em consequência do baixo engajamento na

explicação da variabilidade visto por X_6 na ACP, e pela similaridade elevada entre X_8 (pavimentação) e X_2 (índice de localização), verificada na análise do *Cluster*.

Figura 3 – Gráficos residuais da regressão aprimorada para terrenos

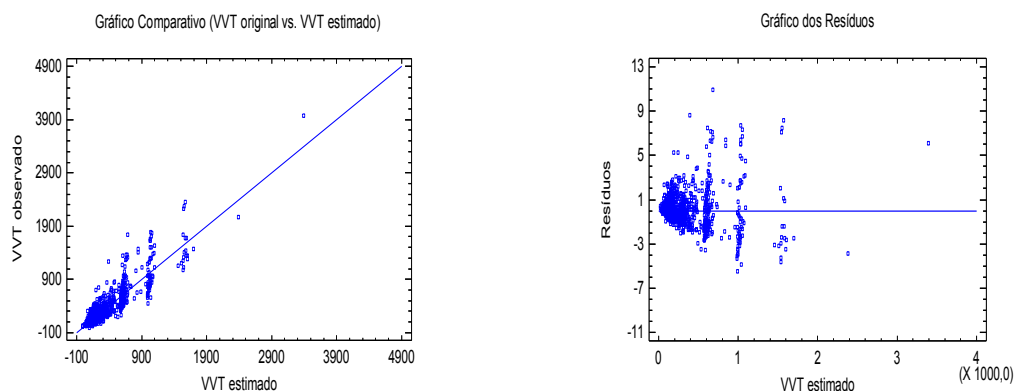
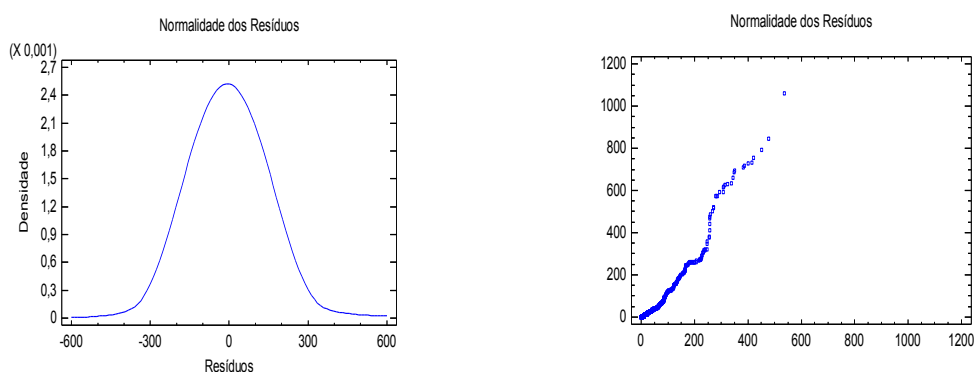


Figura 4 – Normalidade dos resíduos da regressão aprimorada para VVT



Na análise dos resíduos da FIGURA 3 pode-se reparar a diminuição dos erros, obviamente em virtude do tratamento e exclusão dos *outliers*, o que culminou na exclusão de 131 imóveis reduzindo a amostra para 3.091 observações. Ainda, as influências das observações possuidoras de grandes erros nos resultados foram eliminadas do modelo. Pelo gráfico da FIGURA 4 de resíduos *versus* VVT observado nenhum indício de padronização é percebido na disposição dos pontos, possibilitando descartar a presença de correlação serial no modelo. É perceptível, pelo aspecto do mesmo gráfico, que os resíduos não apresentam nenhuma evidência de comportamento tendencioso (crescimento, decrescimento ou oscilação), significando que qualquer indício de heterocedasticidade também é descartado na regressão obtida.

Assim, considerados os ótimos resultados concatenados da Análise de Regressão Linear Múltipla, da ACP e do *Cluster*, julgou-se prudente optar pela adoção da função gerada pelo modelo aperfeiçoado em (2) para realização das estimações práticas pertinentes desta pesquisa, que estão explanadas nas conclusões deste artigo.

5.2 Análise para imóveis edificadas

Buscando-se obter uma função que estime o Valor Venal Predial (VVP), representado por Y_p , através do relacionamento com as variáveis $X_9, X_{11}, \dots, X_{24}$, a exemplo do procedido no caso dos terrenos baldios, inicialmente foi realizada uma análise de regressão englobando todos os 6.813 imóveis contidos na matriz CARACTIMOV EDIFICADOS.

Vislumbrada a hipótese mor de se aprimorar o modelo, em decorrência do surgimento demasiado de *outliers* e algumas correlações significativas entre as variáveis, partiu-se para a efetivação da ACP e *Cluster*.

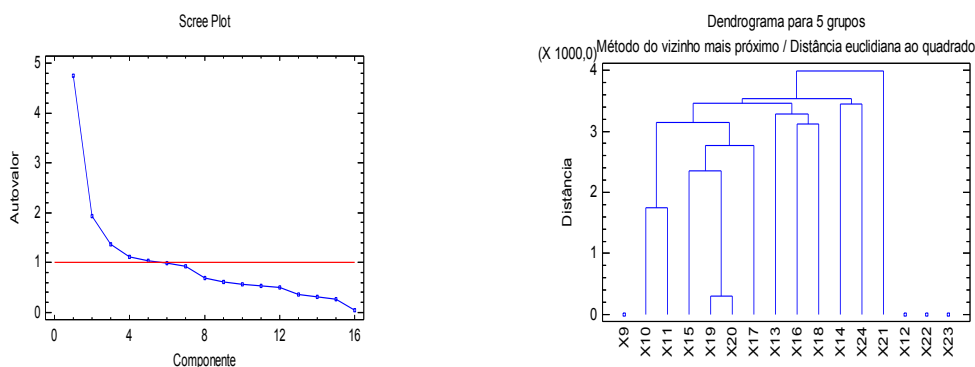
Quadro 3 – Resultados da ACP pelo Critério de Kaiser para edificações

Principal Components Analysis			
Data variables: X9 (ÁREA CONSTRUÍDA), X10 (ACABAMENTO EXTERNO), X11 (ACAB INTERNO), X12(COBERTURA), X13 (PISO), X14 (ESQUADRIAS), X15 (ESTRUTURA), X16 (FORRO), X17(INST ELÉTRICA) X18 (INST SANITÁRIA), X19 (REV EXTERNO), X20 (REV INTERNO), X21(N° DE PAVIMENTOS), X22(LOC HOR), X23(LOC VERT), X24(CONSERVAÇÃO)			
Data input: observations	Standardized: yes	Number of components extracted: 5	
Principal Components Analysis			
Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	4,75357	29,710	29,710
2	1,9411	12,132	41,842
3	1,3643	8,527	50,369
4	1,11196	6,950	57,318
5	1,03642	6,478	63,796
6	0,991246	6,195	69,991
7	0,919296	5,746	75,737
8	0,683024	4,269	80,006
9	0,618111	3,863	83,869
10	0,562668	3,517	87,386
11	0,52606	3,288	90,673
12	0,50424	3,152	93,825
13	0,36061	2,254	96,079
14	0,315045	1,969	98,048
15	0,267129	1,670	99,717
16	0,0452238	0,283	100,000

A ACP está com resultados apresentados no QUADRO 3 e FIGURA 5. Possibilitou concluir que o uso de 5 componentes seria capaz de dirimir 63,79% da variabilidade total presente nos dados, o que não reflete um índice bom. Porém, observou-se que a primeira CP isolada representa quase 30% da variabilidade íntegra existente. Em complementação, a análise de agrupamentos, desempenhada pelo método do vizinho mais próximo, demonstrou que a divisão das variáveis em 5 grupos manteve unidas em um grande conjunto as variáveis $X_{10}, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19}, X_{20}, X_{21}, X_{24}$ pelas suas similaridades, sendo que cada um dos demais 4 grupos detiveram como componentes apenas uma das variáveis faltantes isoladas. Isso pode ser visualizado pelo dendrograma também exposto na FIGURA 5.

O comportamento das variáveis pode ser melhor compreendido após as conclusões retro. Colocadas estas interpretações ainda se faz necessário, para completar a cadeia de verificações e reduzir a interferência residual nas respostas, analisar os *outliers* inerentes no modelo a ser aprimorado. De cada estimação foi extraído o erro inerente, e estes resultados formaram o vetor dos resíduos. Assim foram identificadas as observações responsáveis pelos maiores erros de estimação, que dadas por *outliers* tiveram suas respectivas observações descartadas da amostra trabalhada, resultando na eliminação de 224 imóveis, o que culminou na redução da amostra para 6492 observações.

Figura 5 – Scree Plot da ACP e dendrograma das variáveis para imóveis edificadas



Com os resultados apurados foi possível gerar uma outra função de regressão, eliminando interferências residuais. Esta regressão indicou pouca

contribuição na explicação da variabilidade advinda das variáveis X_{10} (acabamento externo), X_{11} (acabamento interno), X_{17} (instalação elétrica), X_{20} (revestimento interno), X_{21} (nº de pavimentos) e X_{23} (localização vertical). Estas variáveis também foram excluídas da análise com base nos seus p -valores calculados, demonstrando contribuições pouco significativas nas respostas. Logo, com as 6492 observações restantes na amostra e considerando apenas as 10 variáveis dos imóveis não descartadas anteriormente, foi executada a geração de uma outra função de regressão linear múltipla, que é a exposta em (3).

$$Y_p = -1403,68 + 4,6619 \cdot X_9 + 0,877648 \cdot X_{12} + 1,09696 \cdot X_{13} + 1,40884 \cdot X_{14} + 0,932932 \cdot X_{15} + 12,9233 \cdot X_{16} + 5,53463 \cdot X_{18} + 4,41005 \cdot X_{19} + 795,286 \cdot X_{22} + 667,43 \cdot X_{24} \quad (3)$$

Os resultados propiciados demonstram adesão de todas variáveis que compõe o modelo (3), pois ao nível de confiança de 95% pode-se dizer que as dez exercem influência significativa na formação das respostas da função.

Quadro 4 – Resultados da regressão linear múltipla aprimorada para edificações

Multiple Regression - YP											
Dependent variable: YP (VVP EM UFM's)											
Independent variables: X9 (ÁREA CONSTRUÍDA), X12 (COBERTURA), X13 (PISO), X14 (ESQUADRIAS), X15 (ESTRUTURA), X16 (FORRO), X18 (INST. SANITÁRIA), X19 (REV. DAS PAREDES), X22 (LOC. HORIZONTAL), X24 (CONSERVAÇÃO)											
Analysis of Variance											
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value						
Model	2,52673E9	10	2,52673E8	7826,88	0,0000						
Residual	2,09257E8	6482	32282,7								
Total (Corr.)	2,73599E9	6492									
R-squared = 92,3517 percent											
Standard Error of Est. = 179,674											
Mean absolute error = 109,84											
Durbin-Watson statistic = 1,70297 (P=0,0000)											
Correlation matrix for coefficient estimates											
	CONST	X9	X12	X13	X14	X15	X16	X18	X19	X22	X24
CONST	1,0000	0,1447	-0,3970	0,0027	0,0646	-0,0121	-0,3081	-0,0637	0,0483	-0,9130	-0,1523
X9	0,1447	1,0000	-0,1132	-0,0126	0,1239	0,0487	0,1170	-0,3116	-0,0291	-0,1612	-0,1579
X12	-0,3970	-0,1132	1,0000	0,1506	-0,0462	-0,2356	0,1745	0,2006	0,1390	0,2093	-0,0068
X13	0,0027	-0,0126	0,1506	1,0000	-0,0813	-0,0985	-0,2651	-0,0661	0,1644	-0,0570	-0,1558
X14	0,0646	0,1239	-0,0462	-0,0813	1,0000	-0,2311	-0,2926	0,0268	-0,0113	-0,0435	-0,2423
X15	-0,0121	0,0487	-0,2356	-0,0985	-0,2311	1,0000	0,1299	0,0808	-0,5617	-0,0314	-0,1434
X16	-0,3081	0,1170	0,1745	-0,2651	-0,2926	0,1299	1,0000	-0,3192	-0,0845	0,1626	0,0174
X18	-0,0637	-0,3116	0,2006	-0,0661	0,0268	0,0808	-0,3192	1,0000	-0,2209	0,0038	-0,0635
X19	0,0483	-0,0291	0,1390	0,1644	-0,0113	-0,5617	-0,0845	-0,2209	1,0000	0,0047	-0,1549
X22	-0,9130	-0,1612	0,2093	-0,0570	-0,0435	-0,0314	0,1626	0,0038	0,0047	1,0000	-0,0590
X24	-0,1523	-0,1579	-0,0068	-0,1558	-0,2423	-0,1434	0,0174	-0,0635	-0,1549	-0,0590	1,0000

No quadro 4 observa-se que R^2 expressa um excelente índice, apresentando 92,35% de capacidade de explicação da variabilidade total pela função Y_p . Já com respeito à análise de significância, a estatística F calculada resulta em 7.826,88 com p -valor zerado. Isso representa que a hipótese de nulidade dos coeficientes é descartada, e conseqüentemente acolhida a hipótese de existência de relações lineares entre as variáveis independentes com Y_p , logo a validação do modelo sob este aspecto é garantida. Nenhuma correlação alta é percebida entre as variáveis, e isso é desejável, pois exclui a possibilidade de presença de multicolinearidade, atestando o estudo isolado das influências de cada variável independente nas respostas do modelo.

A exclusão dos *outliers* antes mencionada impactou significativamente de forma positiva nos resultados para o modelo (3). Quando se compara o presente resultado ao proveniente da análise inicial de VVP isto fica mais evidente, bem como se verificamos os gráficos oriundos destas análises. Vê-se que as 224 observações descartadas, detentoras dos erros mais discrepantes de outrora, exerciam considerada influência na geração do modelo de regressão, atraindo as respostas da função fora do condizente com as demais observações. Tal influência foi reduzida, satisfazendo o intuito de aprimoramento do modelo.

Figura 6 – Gráficos residuais para a regressão aprimorada para imóveis edificados

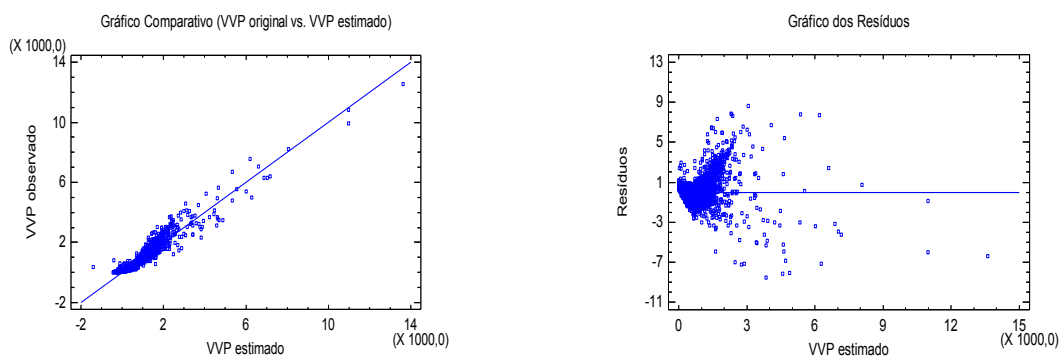
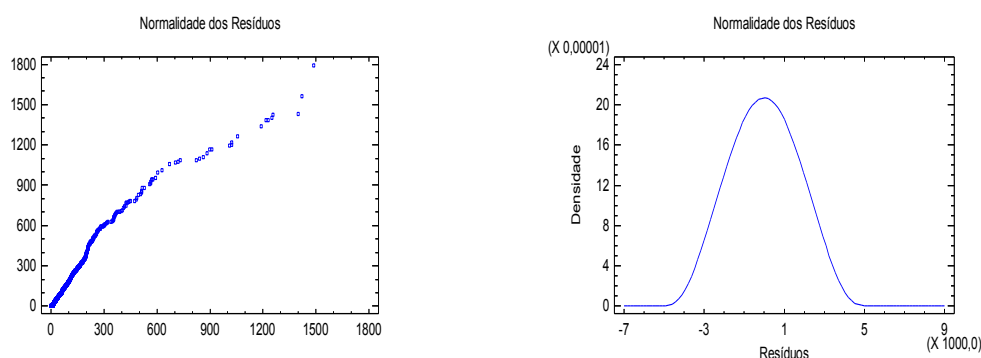


Figura 7 – Normalidade dos resíduos da regressão aprimorada para VVP



Nenhum indício de padrões é refletido pelo gráfico dos resíduos *versus* VVP estimado da FIGURA 6. Este fato aliado ao indicador da estatística de Durbin-Watson corroboram na confirmação de ausência de autocorrelação residual. O gráfico permite ainda a percepção da disposição esparsa dos pontos, sem que detecte-se tendências indesejadas no comportamento dos resíduos conforme FIGURA 7. Logo conclui-se que há variabilidade constante nos erros, o que implica na bem vinda homocedasticidade dos resíduos garantida.

Logo, sendo a normalidade dos resíduos pressuposto inicial existência do modelo múltiplo de regressão, nada se pode alertar para correções no modelo, e a análise residual é concluída, atestando a validade da função obtida em (3).

Encerradas as verificações pode-se concluir que a utilidade da função de regressão múltipla aprimorada pelo modelo (3) logrou êxito após submetida à todos os testes realizados.

Isso posto, combinando os dados advindos da Análise de Regressão Linear Múltipla às interpretações comportamentais das variáveis possibilitadas pela ACP e *Cluster*, entendeu-se por ideal utilizar a função proveniente do modelo aperfeiçoado em (3) para as estimações práticas acerca dos valores venais das construções dos imóveis trabalhados.

6 CONCLUSÕES

A proposta das análises deste artigo é fornecer uma ferramenta de estimação do valor venal de imóveis para fins tributários, com alicerce em técnicas estatísticas multivariadas. Por meio de um estudo de caso, usando dados cadastrais reais

fornecidos pela Prefeitura de um município paranaense, a aplicabilidade dos modelos propostos foi testada, e algumas conclusões podem ser explanadas.

Foram estimados os valores venais de todos os imóveis possuidores de cadastro completo inclusos no banco de dados imobiliário do município estudado, pelas funções em (2) e (3). Sob os valores estimados foram aplicadas as alíquotas vigentes referentes ao IPTU, e assim calculado o imposto incidente caso o valor dos modelos fosse considerado. Ainda o vetor de resíduos foi construído, comportando as diferenças de valores entre o cadastro real da prefeitura com os estimados pelo modelo. Com base nos resultados concluiu-se:

- Foram identificadas variáveis obsoletas utilizadas para as valorizações dos bens, uma vez que os valores dados pelos modelos, obtidos através de um número menor de dados, fornecem resultados consistentes;

- Pelo vetor dos resíduos foi possível constatar que, em esmagadora maioria, os erros não ultrapassaram o índice de 8%, confirmando a excelente performance das estimações;

- Vários dos imóveis tratados por *outliers* nas análises incidiram sobre observações detentoras de grandes áreas do lote, acima de 10 mil m². Talvez tais imóveis mereçam atenção municipal em reavaliações cadastrais e fiscalizações, ou ainda, se localizados em regiões periféricas ao quadro urbano, sejam tratados por rurais, donde não há incidência de IPTU;

- As variáveis que, na prática, exerceram maior influência nos resultados foram índice de localização e área construída, e variações mesmo que leves nos valores destes atributos são os que mais impactaram nas respostas dos modelos;

- Caso os modelos propostos fossem adotados pelo município no exercício de 2009, a arrecadação municipal, provida apenas pelos imóveis abordados nas análises, sofreria acréscimo em torno de 9,89%;

- Possibilidade de utilização contínua do modelo, isso em virtude de sua adaptação às atualizações dos índices de localização (variável X_2) e valores das UFM's;

- Extensão da aplicabilidade do modelo para a taxação do ITBI, com possibilidade de uso para fiscalização e conferência dos valores de venda praticados ou divulgados pelos contribuintes no momento da tributação por venda de imóveis;

É neste enfoque que a pesquisa realizada teve a intenção de oferecer contribuições no estudo da tributação imobiliária, vinculando artifícios úteis também à área da Engenharia de Avaliações, aplicáveis para outros fins, uma vez que os resultados provenientes apresentaram-se consistentes e com precisão satisfatória. Ainda, a partir dos trabalhos aqui apresentados, estudos em outros casos afins podem ser desempenhados, talvez valendo-se do banco de dados cadastrais de outros municípios brasileiros, com possibilidades de aperfeiçoamento nas aplicabilidades, técnicas e resultados.

REFERÊNCIAS

ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2** Avaliação de bens – parte 2: imóveis urbanos. Rio de Janeiro, 2004.

ALVES, Valdir. **Avaliação de imóveis urbanos baseada em métodos estatísticos multivariados**. Dissertação (Mestrado)-vPPGMNE/UFPR. Campo Mourão, 2005.

BAKKE, Hanne A., LEITE, Alexandre S. M., SILVA, Luiz B. Estatística multivariada: aplicação da análise fatorial na engenharia de produção. **Revista Gestão Industrial**, v. 4. Ponta Grossa, 2008.

CÓDIGO TRIBUTÁRIO NACIONAL – **Lei Federal 5172 de 25 de outubro de 1966**

FERREIRA, Daniel F. **Estatística multivariada**. Lavras: Editora UFLA, 2008.

GONZÁLEZ, Marco A. S.. **Aplicação de técnicas de descobrimento de conhecimento em base de dados e de inteligência artificial em avaliação de imóveis**. Tese (Doutorado) - PPGEC/UFRGS. Porto Alegre, 2002.

JOHNSON, Richard A.; WICHERN, Dean W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 1998.

LOPES, Felipe D. L.; SAMOBYL Robert W. Análise de componentes principais: técnica alternativa de análise de confiabilidade em sistemas complexos multivariados. **Revista Produção Online**. Florianópolis, v. 3, jun. 2003.

MATOS, Livia A.; PORTELLA, André A. O aspecto quantitativo do imposto predial e territorial urbano no município de Itauna. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UESC, 11,. 2005. **Anais...UESC**, 2005.

NGUYEN, N; CRIPPS, A. Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. **Journal of Real Estate Research**, Fullerton, California, v. 22, 2001.

PERNAMBUCO, Marcio A. **O CREA e o IPTU**. Disponível em: <<http://www.ibapebahia.org.br/>>. Acesso em: 08 jun. 2010.

PREFEITURA MUNICIPAL DE SÃO MATEUS DO SUL. **Código tributário municipal de São Mateus do Sul**: Lei Complementar 08/04. 2004.

ROMESBURG, Charles H. **Cluster analysis for researchers**. North Carolina: Lulu Press, 2004.

SKIDMORE, Mark; BALLARD, Charles L.; HODGE, Timothy R. Property value assessment growth limits and redistribution of property tax payments: evidence from Michigan. **National Tax Journal**, USA, Set. 2010.

STEINER, Maria T. A.; CHAVES NETO, Anselmo; BAULIO, Silvia N.; ALVES, Valdir. Métodos estatísticos multivariados aplicados à engenharia de avaliações. **Gestão & Produção**. São Carlos, v. 15, jan./abr. 2008.

VICINI, Lorena; SOUZA, Adriano M. **Análise multivariada da teoria à prática**. Santa Maria: UFSM, 2005.



Artigo recebido em 02/02/2012 e aceito para publicação em 19/06/2012.

APÊNDICE I – Variáveis Utilizadas nas Análises com Respectivas Descrições

VARIÁVEIS INDEPENDENTES PARA TERRENOS BALDIOS

VARIÁVEL	CATEGORIAS / UNIDADE DE MEDIDA	DESCRIÇÃO
Área do lote	Metros quadrados	Área total do terreno que constitui o imóvel.
Índice de Localização	Conforme PGV (Planta Genérica de Valores) divulgada anualmente pela Prefeitura Municipal.	Valor utilizado como referência de localização do imóvel. Maiores índices representam localizações mais privilegiadas quanto à proximidade de escolas, bancos, hospitais, áreas comerciais, áreas de lazer, considerando ainda a presença ou ausência de iluminação pública.
Topografia	Por classificação, sendo: 0,7 → irregular; 0,7 → declive; 0,8 → alicive; 1,0 → plano.	Classificação atribuída conforme a situação da superfície da área do terreno do imóvel.
Pedologia	Por classificação, sendo: 0,7 → alagado; 0,8 → inundável; 0,6 → combinação; 1,0 → seco.	Classificação diz respeito às condições do solo do terreno do imóvel.
Situação	Por classificação, sendo: 1,0 → uma frente; 1,1 → duas frentes; 1,5 → três frentes; 2,0 → quatro frentes.	Classificação apresenta o posicionamento do imóvel na quadra em que se localiza.
Fração Ideal	Porcentagem	Traz a quantidade de área utilizável/explorável no terreno do imóvel.
Testada	Metros lineares	Indica a metragem do terreno que faz frente ao logradouro público.
Pavimentação	Por classificação, sendo: 0,5 → terra; 1,0 → par. irregular; 1,5 → par. regular; 2,0 → asfalto.	Classificação atribuída arbitrariamente, levando em consideração o tipo do pavimento do logradouro do imóvel.

VARIÁVEIS INDEPENDENTES PARA IMÓVEIS EDIFICADOS / EM OBRAS

VARIÁVEL	CATEGORIAS / UNIDADE DE MEDIDA	DESCRIÇÃO
Área construída	Metros quadrados	Área total da edificação existente no imóvel.
Tipo	Classificação, assumindo uma das definições: apartamento, casa, sala, loja, galpão, telheiro, indústria, especial.	Indica a espécie de construção existente no imóvel, conforme a classificação em que assume. Foi usada para a definição dos pesos das demais variáveis, porém não figura nos modelos.
Acabamento externo	Classificação numérica *, podendo assumir: sem; caiação; pintura simples; pintura lavável; especial.	Indica qual o tipo de material usado no acabamento do exterior da construção.
Acabamento Interno	Classificação numérica *, podendo assumir: sem; caiação; pintura simples; pintura lavável; especial.	Indica qual o tipo de material usado no acabamento do interior da construção.
Cobertura	Classificação numérica *, podendo assumir: zinco; telha; cimento/amianto; laje; especial.	Apresenta o material utilizado na cobertura/telhado da edificação.
Piso	Classificação numérica *, podendo assumir: terra; taco; assoalho/carpet; cimento; cerâmica.	Refere-se ao material predominante no piso da construção do imóvel.
Esquadrias	Classificação numérica *, podendo assumir: rústica; madeira padrão; ferro; madeira especial; alumínio.	Trata-se do material que compõe as portas e janelas da edificação.
Estrutura	Classificação numérica *, podendo assumir: madeira simples; madeira dupla; mista; alvenaria simples; alvenaria concreto.	Indica a espécie de material predominantemente constante na estrutura (paredes, suporte) do prédio.
Forro	Classificação numérica *, podendo assumir: sem; madeira; estuque; laje; especial.	Definida pelo material utilizado na construção da forragem ante-cobertura existente na edificação.
Instalação elétrica	Classificação numérica *, podendo assumir: sem; até 3 lâmpadas; aparente; semi-embutida; embutida.	Classifica o tipo da instalação elétrica comportada pela construção.
Instalação sanitária	Classificação numérica *, podendo assumir: sem; externa; interna simples; completa; mais de uma.	Refere-se ao(s) banheiro(s) do prédio, quanto ao seu(s) porte(s).
Revestimento externo	Classificação numérica *, podendo assumir: sem; reboco; massa; cerâmica; especial.	Indica qual o tipo de material usado no revestimento das paredes do exterior da construção.
Revestimento interno	Classificação numérica *, podendo assumir: sem; reboco; massa; cerâmica; especial.	Indica qual o tipo de material usado no revestimento das paredes do interior da construção.
Nº de pavimentos	Número natural	Quantia de andares existentes no prédio.

VARIÁVEIS INDEPENDENTES PARA IMÓVEIS EDIFICADOS / EM OBRAS

VARIÁVEL	CATEGORIAS / UNIDADE DE MEDIDA	DESCRIÇÃO
Localização horizontal	Por classificação, sendo: 0,8 → alinhada; 0,7 → recuada; 0,6 → fundos; 0,6 → vila; 0,6 → germinada; 1,0 → superposta.	Diz respeito ao posicionamento da edificação no espaço do terreno/lote.
Localização vertical	Por classificação, sendo: 0,8 → subsolo; 1,0 → térreo e sobreloja; 0,7 → 1º ao 3º andar; 0,7 → 4º ao 6º andar.	Identifica o imóvel quanto sua posição nos pavimentos do prédio.
Conservação	Por classificação, sendo: 1,0 → ótimo; 0,8 → bom; 0,6 → regular; 0,4 → péssimo; 0,2 → em ruínas.	Classifica o estado de conservação da edificação.

* A classificação numérica é condicionada a definição da variável tipo assumida pela edificação, segundo estipula o CTM do município estudado.