

Métodos estatísticos multivariados aplicados à engenharia de avaliações



Maria Teresinha Arns Steiner

Anselmo Chaves Neto

Sílvia Neide Braulio

Valdir Alves

Resumo

O presente artigo tem por finalidade apresentar uma metodologia, composta por técnicas de Análise Multivariada, para a construção de um modelo estatístico de regressão linear múltipla para avaliação de imóveis em função de suas características (variáveis, atributos). É aplicada, inicialmente, a análise de agrupamento aos dados de cada classe de imóvel urbano (apartamentos, residências e terrenos) para obtenção de grupos homogêneos dentro de cada classe e, em correspondência, são determinados discriminantes para alocar futuros itens nesses grupos, pelo método do escore discriminante quadrático. Em seguida, aplica-se a técnica da análise de componentes principais (ACP) para resolver o problema da multicolinearidade que pode existir entre as variáveis do modelo. Os escores das componentes principais são, então, as novas variáveis independentes, e com elas é ajustado um modelo de regressão linear múltipla para cada grupo de imóveis homogêneos dentro de cada classe. A metodologia foi aplicada a um conjunto de 119 imóveis, referentes a 44 apartamentos, 51 casas e 24 terrenos, da cidade de Campo Mourão, PR. O modelo para cada grupo homogêneo, dentro de cada classe de imóveis avaliados, apresentou um ajuste adequado aos dados e uma capacidade preditiva bastante satisfatória.

Palavras-chave: Engenharia de avaliações. Análise de agrupamentos. Análise de componentes principais. Regressão linear múltipla.

1 Introdução

O mercado imobiliário é uma das áreas mais dinâmicas do setor econômico terciário, e as principais dificuldades para a avaliação dos bens advêm das características (atributos, variáveis) dos imóveis que são bastante heterogêneas e podem guardar relacionamentos entre si. A avaliação de imóveis, seja para a cobrança de impostos, venda, garantia em financiamentos ou outros, é feita, em geral, de forma subjetiva com base na experiência pessoal dos gerentes imobiliários e de outros profissionais, que comparam os dados do imóvel da transação com os de imóveis já negociados. Na maioria das vezes não é utilizado qualquer procedimento científico, de forma sistemática, para esta finalidade.

O objetivo do presente artigo é propor uma metodologia baseada em técnicas estatísticas, que seja capaz de fazer previsão do valor de um imóvel levando em consideração os registros históricos de imóveis semelhantes. Esses registros de valores são os definidos em negócios concretizados anteriormente. Para isso, considerou-se

como estudo de caso, o mercado imobiliário da cidade de Campo Mourão, PR, nas classes de apartamentos, casas e terrenos. Deste modo, obtido um modelo estatístico que melhor represente o mercado analisado, em um determinado período de tempo, poder-se-á prever o valor (preço) de mercado de um imóvel qualquer com a máxima precisão possível.

Este artigo está estruturado da seguinte forma: na seção 2 tem-se o enquadramento do problema na área de engenharia de avaliações, com a apresentação das principais normas e conceitos relacionados ao tema; na seção 3 estão alguns trabalhos correlatos ao assunto aqui abordado, e também são descritos os dados para o problema prático em pauta. Na seção 4 são descritas, de forma sucinta, as técnicas estatísticas utilizadas neste trabalho, e também é apresentada a metodologia proposta; na seção 5 tem-se a obtenção e análise dos resultados da aplicação da metodologia proposta aos dados do problema. Finalmente, na seção 6 são apresentadas as conclusões do trabalho.

2 Engenharia de Avaliações

Dantas (2003) define a engenharia de avaliações como uma parte da engenharia que reúne um conjunto de conhecimentos dessa área, da arquitetura e de outras (ciências sociais, exatas e da natureza), com o propósito de determinar, de uma forma técnica, o valor de um bem, de seus direitos, frutos e custos de reprodução, subsidiando tomadas de decisões a respeito de valores envolvendo bens de qualquer natureza. Pode ser praticada por engenheiros, arquitetos e agrônomos; cada um dentro de sua habilitação profissional conforme as leis do Conselho Federal de Engenharia e Arquitetura (CONFEA).

Os primeiros trabalhos publicados no Brasil na área de engenharia de avaliações, que se tem conhecimento, datam do início do século XX. Em 1923, foram introduzidos métodos para avaliação de terrenos que, a partir de 1929, começaram a ter uso sistematizado (FIKER, 1997).

A Associação Brasileira de Normas Técnicas (ABNT) é o Fórum Nacional de Normalização. Em meados de 1950, surgiram as primeiras normas de avaliação de imóveis organizadas por entidades públicas e institutos. O primeiro anteprojeto de normas da ABNT na engenharia de avaliações data de 1957, sendo que a primeira norma brasileira para a avaliação de imóveis urbanos data de 1977, a NB-502/77 (DANTAS, 2003). Esta norma foi revista em 1989, originando a NBR 5676 (ou NB-502/89) registrada no INMETRO.

De acordo com a NB-502/89, os imóveis podem ser classificados quanto ao uso (residencial, comercial, industrial, institucional e misto); quanto à classe de imóvel (terreno, apartamento, casa, escritório, loja, galpão, vaga de garagem, misto, hotéis, hospitais, cinemas e teatros, clubes e recreativos); e quanto ao agrupamento (loteamento, condomínio de casas, prédio de apartamentos, conjunto habitacional, conjunto de salas comerciais, prédio comercial, conjunto de prédios comerciais, conjunto de unidades comerciais, *shopping-centers* e complexo industrial). Vale salientar que o presente trabalho utilizou dados correspondentes apenas aos imóveis das classes apartamentos, casas residenciais e terrenos.

É interessante verificar, que por existirem inúmeras influências para a definição do valor de um imóvel, uma parte do mesmo pode ser considerada aleatória, ou seja, pode-se pensar no valor final com base em um valor mais provável, aumentado ou diminuído de uma parcela imprevisível, de acordo com certas influências pontuais. Deste modo, o valor de um imóvel segue o modelo estatístico: $Y = \mu + \varepsilon$, no qual Y é o valor (preço) negociado; μ é o valor provável e ε é o termo de perturbação estocástica; assim, a esperança de Y é $E(Y) = \mu$. Para um melhor

detalhamento, pode-se consultar, dentre outros, Braulio (2005).

De acordo com González e Formoso (2000), o mercado imobiliário possui um comportamento diferenciado dos mercados de outros bens, devido às características especiais dos imóveis possuem inúmeras fontes de divergências e desigualdades entre os mesmos, impossibilitando assim a comparação direta. Dentre os fatores que diferenciam os imóveis entre si, pode-se citar, dentre outros: a grande vida útil, a fixação espacial, a singularidade, o elevado prazo de maturação e o alto custo das unidades.

A ABNT (NBR5676/90) divide os métodos de avaliação em dois grandes grupos: métodos diretos e indiretos. Um método é considerado direto quando o valor do resultado da avaliação independe de outros (DANTAS, 2003). Os métodos diretos subdividem-se em método comparativo de dados de mercado (define o valor através da observação comparativa de valores com outros itens semelhantes no mercado) e método comparativo de custo de reprodução de benfeitorias (apropria o valor das benfeitorias). Segundo Dantas (2003), há preferência pelo emprego dos métodos diretos, e sempre que existem dados de mercado suficientes para a sua utilização, são os escolhidos.

Já um método é considerado indireto quando necessita de resultados de algum método direto. Os métodos indiretos subdividem-se em método da renda (define o valor em função de um rendimento já existente ou previsto pelo bem no mercado, ou seja, pelo valor econômico do bem); método involutivo (o valor é estimado por estudos de viabilidade técnica-econômica do seu aproveitamento) e método residual (calcula a diferença entre o valor total do imóvel e o valor das benfeitorias, levando-se em conta o fator de comercialização).

Quanto aos níveis de precisão, os trabalhos de avaliação podem ser classificados da seguinte forma: de nível de rigor expedito (o valor é obtido sem a utilização de qualquer instrumento matemático); normal (utilizam-se métodos estatísticos e existem exigências com relação à coleta e ao tratamento dos dados); rigoroso (o valor, resultado do tratamento estatístico adotado, deverá ter um nível de confiança máximo de 80%, com as hipóteses nulas testadas a um nível de significância máximo de 5%); e rigoroso especial, que se caracteriza pelo encontro de um modelo estatístico o mais abrangente possível, ou seja, que incorpore o maior número de características que contribuem para a formação do valor.

A função estimada para a formação do valor deve ser eficiente e não tendenciosa. A hipótese nula sobre o modelo de regressão deve ser rejeitada somente com nível de significância máximo de 1% (ANOVA). Já as hipóteses nulas sobre os parâmetros do modelo de regressão devem ser testadas com nível de significância máximo de 10%

para o teste unilateral (teste t) ou 5% em cada ramo no teste bilateral. Devem ser analisadas as seguintes condições básicas referentes aos resíduos do modelo ajustado aos dados: Gaussianidade, homogeneidade da variância e independência. Desta forma os resíduos devem ser Gaussianos, independentes e identicamente distribuídos, ou seja, $\varepsilon_i \sim N(0, \sigma^2)$.

3 Trabalhos correlatos e descrição do problema

Existem alguns trabalhos na literatura que tratam da Engenharia de Avaliações. Pode-se citar, por exemplo, o trabalho de Nguyen e Cripps (2001), em que comparam o desempenho preditivo de Redes Neurais Artificiais com a Análise de Regressão Múltipla para a venda de casas de família. Diversas comparações foram feitas entre os dois modelos, nas quais foram variados: o tamanho da amostra de dados, a especificação funcional e a predição temporal. Já no trabalho de Bond et al. (2002), os autores examinam o efeito que a vista de um lago (Lago Erie, E.U.A.) tem sobre o valor de uma casa. No estudo, foram levados em consideração os preços baseados na transação das casas (preço de mercado). Os resultados indicam que além da variável vista, a qual se apresenta significativamente mais importante que as demais, também a área construída e o tamanho do lote são importantes.

Cechin et al. (1999) compararam a técnica de regressão linear e redes neurais artificiais para realizar a estimativa do custo de venda e de aluguel de imóveis em Porto Alegre, RS. Foram avaliados dois bancos de dados: 1.600 imóveis ofertados para venda com 20 atributos cada e 500 imóveis ofertados para aluguel com 85 atributos cada. Do total destes atributos, apenas seis foram selecionados para treinamento dos modelos. Guedes (2001) apresenta duas ferramentas ao engenheiro de avaliações: modelos lineares generalizados e Redes Neurais aplicadas a 50 lotes urbanos de três bairros da cidade de Recife, PE.

Pelli Neto e Zárate (2003) também fizeram um estudo comparativo do uso de redes neurais e análise de regressão múltipla para estimação do valor de venda de imóveis, referente à oferta de 172 apartamentos de média e baixa renda do mercado imobiliário de Belo Horizonte, MG. Brondino (1999) apresentou um trabalho utilizando redes neurais na determinação da influência da variável acessibilidade no valor de lotes urbanos, comparando-as com o modelo de regressão múltipla, em duas cidades do interior de São Paulo (São Carlos e Araçariçuama). A referida variável apresentou um peso superior a 34% no preço final do imóvel.

A aplicação da metodologia aqui proposta foi aplicada aos imóveis urbanos nas classes de apartamentos, residências e terrenos da cidade de Campo Mourão, PR. A

amostra foi constituída por 119 imóveis (classes), sendo 44 da classe de apartamentos, 51 da classe de residências e 24 da classe de terrenos. Todos estão localizados na área urbana da cidade; destes, 80 estão localizados na área central da cidade.

Os atributos são dos tipos qualitativo e quantitativo; os de apartamentos estão listados no Apêndice 1 que, como se pode observar, são 21 ao todo, já divididos em grupos (seções posteriores), sendo 17 do grupo A, 19 do grupo B e 19 do grupo C.

4 Metodologia aplicada

A metodologia para a obtenção do objetivo consta da aplicação das seguintes técnicas estatísticas da área de análise multivariada:

Análise de agrupamento: através desta técnica procura-se determinar para cada classe de imóveis, os grupos de itens homogêneos. Nesta análise utilizou-se a Distância Euclidiana, e o método de ligação foi o método de Ward.

Após a formação dos grupos homogêneos, foram construídas funções discriminantes com duas finalidades: avaliar a consistência dos grupos obtidos; e alocar futuros itens em cada um dos grupos que formam cada classe.

Em seguida, aplicou-se a análise de componentes principais para cada um dos grupos, de cada uma das classes, com o objetivo de substituir os valores das variáveis originais pelos escores das componentes principais e contornar o possível problema da multicolinearidade.

Finalmente, foi ajustado um modelo de regressão linear múltipla para cada um dos grupos de cada classe de imóvel. O preço à vista, denominado valor, foi considerado como variável resposta ao modelo.

4.1 Descrição das técnicas estatísticas utilizadas

A Análise Multivariada é um conjunto de técnicas usadas para resolver problemas relacionados a: 1) Estrutura de covariância do vetor aleatório \underline{X} (resumida na matriz de covariância ou de correlação) através da análise de componentes principais; análise fatorial e análise de correlação canônica; 2) Agrupamento de itens (*cluster analysis*); 3) Reconhecimento e Classificação de Padrões (JOHNSON; WICHERN, 1998). Nesta seção são descritas, de forma sucinta, as técnicas estatísticas multivariadas utilizadas.

4.1.1 Análise de agrupamentos

A Análise de Agrupamento consiste em uma técnica que tem o objetivo de formar grupos homogêneos de objetos (imóveis). Os grupos são formados com base nas suas distâncias (Euclidiana, Mahalanobis, dentre outras) ou similaridades e em um método de ligação entre os grupos

parciais. A distância usualmente utilizada é a Distância

Euclidiana: $d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$. O método de

ligação mais utilizado é o de Ward, que minimiza a “perda das informações” ao “juntar” dois grupos, utilizando o critério de minimizar a soma dos quadrados dos erros

(*sum quadratic error, SQE*), $SQE = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})$.

4.1.2 Escore de discriminação quadrático para reconhecimento e classificação

Neste estudo usou-se a regra de reconhecimento e classificação com base na probabilidade total mínima de erro definida pelo escore de discriminação quadrático para a população (grupo) i dado por:

$$d_i^Q(\underline{x}) = \quad (1)$$

$$-\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\underline{x} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) + \ln(p_i) \quad i = 1, 2, \dots, g$$

em que p_i é a probabilidade, a priori, do item pertencer à população Π_i , já $\underline{\mu}_i$ e Σ_i são, respectivamente: o vetor médio e a matriz de covariância da população i . Estes parâmetros são geralmente desconhecidos e, assim sendo, trabalha-se com as suas estimativas amostrais \bar{x}_i e S_i . Quanto as p_i , pode-se tomá-las como as proporções dos tamanhos dos grupos amostrais. Reconhece-se \underline{x}_0 como sendo de Π_k se: $d_k^Q(\underline{x}_0) > d_i^Q(\underline{x}_0) \quad \forall i = 1, 2, \dots, g$, com $k \neq i$.

4.1.3 Análise de componentes principais

Seja o vetor aleatório x com p componentes correlacionadas. A estrutura desse relacionamento pode ser resumida na matriz de covariância Σ ou na matriz de correlação ρ . Sabe-se do Teorema da Decomposição Espectral, que $\Sigma = PAP'$ ou $\rho = PAP'$, em que P é a matriz ortogonal dos autovetores e Λ a matriz diagonal dos autovalores. Assim, existem p componentes principais não correlacionadas representadas pelas combinações lineares $Y_i = e_i'X$, que recompõem essa estrutura de covariância, em que e_i e λ_i , $i = 1, 2, \dots, p$ são, respectivamente, os autovetores e autovalores de Σ ou ρ .

Além disso, é bem conhecido que $V(Y_i) = \lambda_i$ exprime a importância de cada componente principal. Um número $m < p$ de componentes principais pode representar uma parcela significativa da variação total e é possível usá-las no lugar das p variáveis originais. Um critério para a determinação do número de componentes principais a serem consideradas foi sugerido por Kaiser, em 1960 (JOHNSON; WICHERN, 1998). Consiste em se tomar um número m de componentes principais igual ao número

de autovalores $\lambda_i \geq 1$. Ainda, é interessante considerar a parcela da variação explicada pelas m componentes principais acima de (cerca de) 90%, ou seja, estendendo o referido critério, podem ser considerados autovalores menores que 1, desde que próximos a 1.

Ao aplicar-se a análise de componentes principais, obtêm-se os escores de suas m principais componentes. Desta forma, transforma-se a matriz X do modelo da ordem $n \times p$ na matriz E de ordem $n \times m$, $m < p$, correspondente aos escores das m componentes principais.

4.1.4 Regressão linear múltipla para previsão

Para se obter o valor de uma variável Y em função de outras variáveis X_i independentes entre si, usa-se um modelo de Regressão Linear Múltipla, dado por: $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$; em que \underline{Y} é o vetor das respostas observadas das n observações (imóveis); X é a matriz do modelo de ordem $n \times p$; ε é o vetor dos erros de dimensão n e $\underline{\beta}$ (a ser estimado) é o vetor de parâmetros de dimensão p .

Uma vez definido o grupo k da classe ℓ do item (imóvel) com base no agrupamento e no reconhecimento e classificação, usa-se o modelo de regressão linear múltipla ajustado para estimar o valor do imóvel j por:

$$\hat{y}_j = \underline{\text{esc}}_j' \hat{\underline{\beta}} \quad (2)$$

em que $\underline{\text{esc}}_j$ é o vetor dos escores das componentes e $\hat{\underline{\beta}}$ é o vetor de parâmetros estimado.

5 Obtenção dos resultados

Os grupos da classe de apartamentos foram formados pela Análise de Agrupamentos descrita no item a da seção 4.1 anterior. O resultado indicou que três grupos compõem a classe dos apartamentos, como mostra a Figura 1. O grupo 1 contém 38,64% dos apartamentos analisados, o grupo 2, 22,73% e o grupo 3, 38,64%.

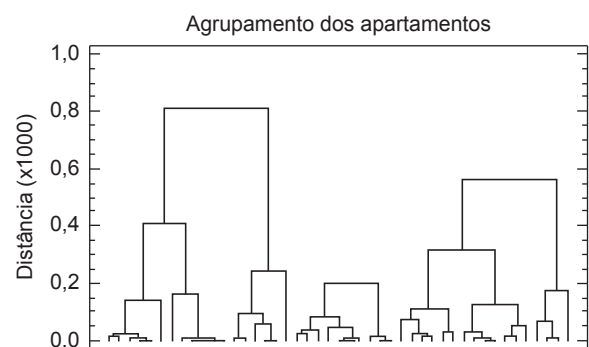


Figura 1. Dendrograma das três classes formadas com os 44 apartamentos.

Após a análise de agrupamento, realizou-se a análise discriminante usando-se os escores quadráticos, sendo que os resultados encontram-se na Tabela 1.

Na Tabela 1 tem-se que, dos 17 apartamentos pertencentes à classe 1, todos foram classificados corretamente; dos 10 pertencentes à classe 2, todos também foram classificados corretamente; e dos 17 pertencentes à classe 3, apenas um foi classificado incorretamente. Deste modo, tem-se a precisão de 43/44. $100 = 97,73\%$. Assim, os resultados apresentaram-se consistentes: dentre as 44 observações que ajustam o modelo, 97,73% foram classificadas corretamente. A interpretação dos grupos obtidos foi realizada de acordo com os atributos de cada classe, sendo 17 do grupo A, 19 do grupo B e 19 do grupo C, conforme já mencionado.

A seguir, são apresentados os aspectos mais determinantes de cada um dos três grupos:

- Grupo 1:** todos os apartamentos estão localizados no centro; estão em prédios com pelo menos sete pavimentos; todos têm pelo menos uma vaga na garagem; o revestimento dos prédios é pastilhado; possuem área mínima de 160 m²; possuem pelo menos um elevador; têm mais de dois dormitórios; todos possuem suíte; todos possuem dependência de empregada completa; e o seu preço é superior a R\$115.000,00.
- Grupo 2:** todos localizados no centro; estão em prédios com pelo menos 13 pavimentos; o revestimento dos prédios é de pastilha, mármore ou granito; todos têm mais de uma vaga na garagem; possuem área mínima de 220 m²; são prédios com menos de 15 anos; possuem mais de dois dormitórios; todos possuem suíte; todos possuem mais de um elevador; todos têm dependência de empregada completa; e o seu preço é superior a R\$175.000,00.
- Grupo 3:** todos os apartamentos possuem apenas uma vaga na garagem; têm área privativa menor que 132 m²; são prédios baixos; e seus valores estão entre R\$30.000,00 e R\$110.000,00.

Em seguida, foi aplicada a Análise de Componentes Principais aos dados das variáveis explicativas originais, obtendo-se $m = 6$ componentes e os seus escores, conforme mostra a Tabela 2.

Tabela 1. Classificação dos 44 apartamentos em três grupos.

Número de classes	Quantidade de imóveis	Classe 1	Classe 2	Classe 3
1	17	17 (100,00%)	0 (0,00%)	0 (0,00%)
2	10	0 (0,00%)	10 (100,00%)	0 (0,00%)
3	17	0 (0,00%)	1 (5,88%)	16 (94,12%)

Através dos resultados contidos na Tabela 3 a seguir, pode-se observar que a primeira componente possui pesos mais altos nas variáveis originais: distância à escola; distância ao supermercado; distância ao hospital; nível de conservação; e número de banheiros. A segunda componente possui pesos mais altos nas variáveis: número de elevadores; número de dormitórios; e idade aparente do imóvel. A terceira componente possui pesos maiores nas variáveis: número de peças do imóvel; número de salas; número de banheiros; e padrão de acabamento. A quarta componente possui pesos mais altos nas variáveis: vaga de carro na garagem; área útil; número de pavimentos; número de salas; padrão de acabamento; e idade real. A quinta componente possui pesos mais altos: posição do apartamento no prédio; número de elevadores; número de vagas na garagem; e número de pavimentos no prédio. Finalmente, a sexta componente possui pesos mais elevados nas variáveis: posição do apartamento no prédio; número de vagas na garagem; e nível de conservação.

Os escores fornecidos pelas seis componentes para os 17 apartamentos compõem a Tabela 4, adiante. Esses são os valores das variáveis explicativas considerados para o ajuste do modelo de regressão linear.

No ajuste do modelo de regressão linear múltipla $\hat{Y} = X\beta + \varepsilon$ verificou-se que a quinta e a sexta componente não são significativamente importantes devido aos seus valores-p serem maiores do que 0,05, sendo assim foram descartadas, e consideraram-se apenas as quatro primeiras, como mostra a Tabela 5.

O coeficiente de correlação múltipla ao quadrado R^2 , que mede a qualidade do ajuste é dado por:

Tabela 2. Análise de componentes principais dos apartamentos pertencentes ao grupo 1 ($m = 6$).

Componente	Autovalor	Variância (%)	% Acumulada
1	4,38168	25,775	25,775
2	3,68641	21,685	47,459
3	2,5373	14,925	62,385
4	1,93819	11,401	73,786
5	1,3989	8,229	82,015
6	0,956367	5,626	87,640
7	0,644585	3,792	91,432
8	0,548348	3,226	94,658
9	0,397581	2,339	96,996
10	0,375483	2,209	99,205
11	0,118514	0,697	99,902
12	0,0117594	0,069	99,971
13	0,00487836	0,029	100,000
14	1,18656E-16	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000
17	0,0	0,000	100,000

Tabela 3. Pesos das variáveis originais em cada uma das seis componentes principais dos apartamentos pertencentes à classe 1.

Variável	Componente					
	1	2	3	4	5	6
ppredio	0,0893909	0,246161	-0,0526886	-0,0979668	0,416273	-0,570934
elevador	-0,0526883	-0,413666	0,0726574	-0,152862	0,35089	-0,0682263
vgaragem	0,0292966	-0,115928	-0,142546	-0,344625	0,513443	0,376004
área const.	0,289871	-0,133004	0,088613	-0,431775	-0,000990526	-0,0850487
pavimento	0,133738	-0,0357628	0,243017	0,47084	0,391287	0,066217
andar	0,208736	0,285383	0,0221086	0,102197	0,246664	-0,27849
peças	0,225747	0,063101	-0,531922	-0,0212474	0,0193336	0,0672196
sala	0,223168	-0,107456	-0,355162	0,311851	-0,0248622	0,274322
quarto	0,127221	0,400858	-0,25558	-0,00234591	-0,1004	0,11388
banheiro	0,308159	-0,0270833	-0,397666	0,147821	0,00518783	-0,255486
descola	-0,343604	0,278393	-0,136115	-0,045763	0,233369	0,169827
dhospital	-0,363872	0,279642	-0,118547	-0,0527298	-0,0264049	0,15829
dsmercado	-0,381969	0,262674	-0,0967804	-0,0107717	0,212156	-0,0572527
acabamento	0,0963664	0,245521	0,347094	0,306414	-0,0709979	0,0243955
conservação	0,327714	0,14195	0,174912	0,0723854	0,268025	0,455252
idreal	0,263027	0,240884	0,114627	-0,410822	-0,194047	0,064233
idaparente	0,220789	0,338124	0,255992	-0,193006	-0,0148376	0,108

Tabela 4. Escores das seis componentes principais dos apartamentos pertencentes ao grupo 1.

Imóvel	Componente					
	1	2	3	4	5	6
1	-2,1512	-0,741581	1,27721	0,867893	-0,379761	-2,77916
2	-0,215068	0,281233	3,1519	1,28665	2,81409	-0,021318
3	0,554301	0,0739132	-0,793701	2,64601	-1,49977	0,403009
4	-0,896709	0,0629922	-0,31083	0,478471	1,86133	0,619036
5	2,97263	-1,50975	2,21074	0,225224	-0,129701	0,859525
6	2,93197	-1,36201	-3,42417	-1,01255	1,87297	0,0124523
7	-0,142195	1,89695	-0,454084	-1,25077	0,675696	0,381225
8	4,01045	-1,83611	-0,735231	-0,644179	-0,672452	-1,73787
9	0,663582	2,92859	-1,21238	1,96973	0,00939094	-0,404564
10	0,28557	2,2559	-1,17425	1,96289	-0,827946	0,66776
11	0,115707	3,10054	0,897485	-2,17611	-0,685172	-
12	0,115707	3,10054	0,897485	-2,17611	-0,685172	-
13	-2,30919	-1,28958	-0,548849	-0,223296	-0,267799	0,384237
14	-2,46238	-1,34315	-0,643239	-0,47653	-0,0581804	0,0206226
15	-2,61556	-1,39672	-0,737628	-0,729764	0,151438	-0,342991
16	-2,84039	-2,01583	-0,605112	-0,48337	-0,895518	1,09295
17	1,98277	-2,20593	2,20466	-0,264193	-1,28344	1,02096

Tabela 5. Ajuste do modelo de regressão linear múltipla para os apartamentos pertencentes à classe 1.

Parâmetro	Estimativa	Erro padronizado	Estatística T	Valor-p
CONSTANTE	174412,0	3150,0	55,3689	0,0000
PCOMP_1	18607,3	1551,15	11,9958	0,0000
PCOMP_2	4386,74	1691,11	2,594	0,0235
PCOMP_3	7100,19	2038,4	3,48323	0,0045
PCOMP_4	-23492,7	2332,26	-10,0729	0,0000

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \text{ com } 0 < R^2 < 1 \quad (3)$$

e fornece o valor de $R^2 = 0,956557$, ou seja, o modelo ajustado explica cerca de 96% da variabilidade do preço de mercado. Assim, a equação de regressão linear múltipla para os apartamentos pertencentes à classe 1, que descreve a relação entre o preço e as quatro componentes independentes é dada pela Equação 4 a seguir.

$$\text{Preço} = 17412,0 + 18607,3 Y_1 + 4386,74 Y_2 + 7100,19 Y_3 - 23492,7 Y_4 \quad (4)$$

Já a análise de variância, contida na Tabela 6, mostra que se deve rejeitar a hipótese de não haver regressão, ou seja, o modelo acima é de fato significativo.

As premissas necessárias ao uso do modelo linear e aos testes aplicados foram todas verificadas e satisfeitas pelos resíduos, ou seja, $\varepsilon_i \sim N(0, \sigma^2)$. Assim, no modelo para os apartamentos, comprovou-se a Gaussianidade dos resíduos pelo teste de Kolmogorov-Smirnov, que forneceu um valor-p de $p = 0,987074$ e a homogeneidade da variância dos resíduos foi verificada e aceita por método gráfico (resíduos x valores preditos).

Os valores preditos pela Equação 4 ajustada e os valores observados, bem como as percentagens de erros da predição são apresentados na Tabela 7.

A análise conduzida para os 10 apartamentos pertencentes à classe 2, resultou em seis componentes principais que explicam 92,44% da variabilidade dos dados originais. E aos escores que compõem a matriz do modelo de ordem (10 x 6) foi ajustado o modelo de regressão linear múltipla $\underline{Y} = X \underline{\beta} + \underline{\varepsilon}$, resultando em:

$$\text{Preço} = 133036,0 + 18814,5 Y_1 + 2245,63 Y_2 - 1733,73 Y_3 + 11164,2 Y_4 - 515,17 Y_5 - 2907,38 Y_6 \quad (5)$$

O coeficiente de determinação do ajuste foi de $R^2 = 0,998942$, ou seja, o modelo ajustado explica quase 100% da variabilidade do preço de mercado. A hipótese da existência do modelo foi testada e aceita pela análise da variância, que forneceu valor-p, no teste F, de

Tabela 6. Análise de variância do ajuste do modelo de regressão para os apartamentos pertencentes ao grupo 1.

Fonte	Soma dos quadrados	Grau de liberdade	Quadrado médio	F	valor-p
Model	4,45699E10	4	1,11425E10	66,06	0,0000
Residual	2,02419E9	12	1,68682E8		
Total	4,65941E10	16			

$p = 0,0002$. Da mesma forma que no grupo 1, as premissas para validação do modelo linear e dos testes aplicados foram todas verificadas e satisfeitas pelos resíduos, ou seja, $\varepsilon_i \sim N(0, \sigma^2)$. Assim, comprovou-se a Gaussianidade dos resíduos pelo teste de Kolmogorov-Smirnov com $p = 0,889229$ e a homogeneidade da variância dos resíduos também foi aceita por método gráfico (resíduos x valores preditos).

Já para os 17 apartamentos pertencentes ao grupo 3, a Análise de Componentes Principais apontou que as sete primeiras componentes explicam 88,949% da variabilidade dos dados originais. O ajuste do modelo $\underline{Y} = X \underline{\beta} + \underline{\varepsilon}$ à matriz dos escores de ordem (17 x 7) resultou em:

$$\text{Preço} = 70705,90 + 3470,01 Y_1 + 6673,20 Y_2 - 1917,89 Y_3 + 4096,52 Y_4 - 7571,59 Y_5 - 6981,03 Y_6 - 7104,54 Y_7 \quad (6)$$

O coeficiente de determinação obtido foi de $R^2 = 0,893306$ e, conseqüentemente, o modelo ajustado explica cerca de 89% da variabilidade do preço de mercado. A existência do modelo foi testada e aceita no teste F da análise de variância, que forneceu $p = 0,0010$. As premissas para validação do modelo linear e dos testes aplicados foram todas verificadas e satisfeitas pelos resíduos, ou seja, $\varepsilon_i \sim N(0, \sigma^2)$. Comprovou-se a Gaussianidade pelo teste de Kolmogorov-Smirnov com $p = 0,110277$ e a homogeneidade da variância foi aceita por método gráfico (resíduos x valores preditos).

Tabela 7. Resultados para os 17 apartamentos do grupo 1.

Valor observado y_i (R\$)	Valor predito \hat{y}_i (R\$)	Erro absoluto $y_i - \hat{y}_i$ (R\$)	Erro percentual $(y_i - \hat{y}_i) (%)$
130.000,00	128.715,00	1.285,00	0,98846
150.000,00	149.891,00	109,00	0,07267
120.000,00	117.273,00	2.727,00	2,2725
170.000,00	170.992,00	992,00	0,583529
250.000,00	244.826,00	5.174,00	2,0696
220.000,00	219.705,00	295,00	0,13409
200.000,00	198.514,00	1.486,00	0,743
250.000,00	249.772,00	228,00	0,0912
150.000,00	146.357,00	3.643,00	2,42867
120.000,00	127.476,00	7.476,00	6,23
250.000,00	249.481,00	519,00	0,2076
250.000,00	249.481,00	519,00	0,2076
115.000,00	114.222,00	778,00	0,67652
120.000,00	128.543,00	8.543,00	7,119167
140.000,00	142.864,00	2.864,00	2,045714
120.000,00	109.661,00	10.339,00	8,61583
210.000,00	217.228,00	7.228,00	3,441905

6 Conclusões

Neste artigo é proposta uma metodologia baseada em técnicas estatísticas multivariadas para a previsão de preços de imóveis urbanos. A referida metodologia é composta das seguintes técnicas: Análise de Agrupamento, na qual os imóveis “semelhantes” em termos de atributos são agrupados; Escores Discriminantes Quadráticos, em que é verificada a consistência dos grupos formados e tem-se um critério para alocação de um novo item. Em seguida, é aplicada a Análise de Componentes Principais para conseguir $m < p$ componentes e os seus escores independentes para substituir as p variáveis originais, contornando-se assim o problema da multicolinearidade. Finalmente, faz-se o ajustamento do modelo de Regressão Linear Múltipla do vetor de valores \underline{Y} contra as variáveis explicativas resumidas na matriz E de ordem $(n \times m)$, ou seja, $\underline{Y} = E\underline{\beta} + \underline{\varepsilon}$, que fornece uma estimativa do valor do imóvel \underline{x}_0 através da Equação 7:

$$\hat{y}_0 = \hat{\beta}' \underline{e}_0 \quad (7)$$

em que \underline{e}_0 é o vetor de escores correspondentes.

A metodologia foi aplicada às outras duas classes de imóveis (51 residências e 24 terrenos) com resultados considerados bastante satisfatórios. A qualidade do ajuste às variáveis, agora de fato independentes, forneceu os coeficientes de determinação apresentados na Tabela 8.

Assim, dado um novo imóvel (apartamento, residência ou terreno) da cidade de Campo Mourão, PR, do qual se

Tabela 8. Valores de R^2 para os grupos das três classes.

Classe	Grupos			
	1	2	3	4
Apartamentos	0,956557	0,998942	0,893306	-
Residências	0,91937	0,99226	0,955551	0,969181
Terrenos	0,977556	0,997455	-	-

deseja ter uma estimativa do seu valor, deve-se inicialmente verificar a que grupo o mesmo pertence, aplicando os escores quadráticos. Identificado o grupo, pode-se utilizar o modelo de Regressão Linear Múltipla correspondente àquele grupo. Para a classe dos apartamentos, grupo 1, o modelo definido é apresentado na Equação 4, seção 5. Da mesma forma, têm-se os modelos para as demais situações. A metodologia é genérica, podendo ser utilizada a qualquer cidade, obtendo-se a definição dos modelos para as diversas situações em cada uma delas.

A metodologia multivariada apresentada para predições dos preços de imóveis se mostrou viável e altamente apropriada, atingindo-se resultados com níveis de precisão bastante satisfatórios. Desta forma, a mesma poderá servir de apoio tanto a gerentes imobiliários na definição de preços de imóveis, como para pessoas físicas e/ou jurídicas que desejam conhecer mais realisticamente o seu patrimônio. Deve-se atentar para o fato de que os modelos de Regressão Linear Múltipla definidos deverão sofrer reajustes periódicos, devido à alta dinamicidade de crescimento e da economia do país.

Multivariate statistical methods applied to evaluation engineering

Abstract

The purpose of this paper is to present a methodology composed by Multivariate Statistical Analysis techniques in order to build a Statistical Multiple Linear Regression Model to evaluate some estates according to their characteristics (variables, attributes). First, a Clustering Analysis was applied to the data of each urban estate class (apartments, houses or lots) to categorize similar groups and, correspondingly, the discriminant was defined in order to assign future items to these groups, by means of the Quadratic Score Discriminant Method. Next, the Principal Components Analysis (P.C.A.) was applied to solve the multicollinearity problem that may exist among the variables in the model. The scores of the principal components become then the new independent variables and with them a Multiple Linear Regression model was adjusted to each group of similar estate within each class. This methodology was applied to a set of 119 estates including 44 apartments, 51 houses and 24 lots in the city of Campo Mourão, PR. The model for each similar group in each class of the evaluated estates presented an adequate adjustment to the data and a satisfactory predictive capacity.

Keywords: Evaluation engineering. Clustering analysis. Principal components analysis. Multiple linear regression.

Referências bibliográficas

- ABNT (Associação Brasileira de Normas Técnicas). **Avaliação de Imóveis Urbanos** (NBR 5676 e NBR 502). Rio de Janeiro: ABNT, 2004.
- BOND, M. T.; SEILER, V. L.; SEILER, M. J. Residential Real Estate Prices: a Room with a View. **The Journal of Real Estate Research**, California State University, Fullerton, California, v. 23, n. 1, p. 129-137, 2002.
- BRAÚLIO, S. N. **Proposta de uma Metodologia para a Avaliação de Imóveis Urbanos baseada em Métodos Estatísticos Multivariados**. Curitiba – PR, 2005, 158 p. Dissertação – (Mestrado em Métodos Numéricos em Engenharia). Programação Matemática. UFPR.
- BRONDINO, N. C. M. **Estudo da Influência da Acessibilidade no Valor de Lotes Urbanos através do uso de Redes Neurais**. São Carlos – SP, 1999, 158 p. Tese – (Doutorado em Engenharia Civil). Transportes – USP.
- CECHIN, A. L.; SOUTO, A.; GONZÁLEZ, M. A. Análise de Imóveis através de Redes Neurais Artificiais na Cidade de Porto Alegre. **Scientia**, São Leopoldo, RS, v.10, n. 2, p. 5-32, 1999.
- DANTAS, R. A. **Engenharia de Avaliações: uma Introdução à Metodologia Científica**. São Paulo: Pini, 2003. 262 p.
- FIKER, J. **Avaliação de Imóveis Urbanos**. São Paulo: Pini, 1997. 104 p.
- GONZÁLEZ, M. A. S.; FORMOSO, C. T. Análise conceitual das dificuldades na determinação de modelos de formação de preços através de análise de regressão. **Engenharia Civil – UM (Universidade do Milho)**, Portugal, n. 8, p. 65-75, 2000.
- GUEDES, J. C. Duas Ferramentas Poderosas a Disposição do Engenheiro de Avaliações: Modelos Lineares Generalizados e Redes Neurais. In: Anais do XI COBREAP, 2001, Guarapari, ES. **Anais...** Guarapari, 2001, 18 p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 1998. 816 p.
- NGUYEN, N.; CRIPPS, A. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. **The Journal of Real Estate Research**, Fullerton, California, v. 22, n. 3, p. 313-336, 2001.
- PELLI NETO, A.; ZÁRATE, L. E. Avaliação de Imóveis Urbanos com a utilização de Redes Neurais Artificiais. In: Anais do IBAPE – XII COBREAP, 2003, Belo Horizonte, MG. **Anais...** Belo Horizonte, MG, 2003. 14 p.

Apêndice 1: Relação dos atributos para apartamentos (fonte: Imobiliária Tapowik, Guarapuava, PR).

Atributos	Descrição	Categorias	Aptos. Grupo A	Aptos. Grupo B	Aptos. Grupo C
ppedio	Identifica o revestimento do prédio.	1 = pintura 2 = pastilha 3 = cerâmica 4 = mármore/granito		X	X
Andar	Escore referente ao andar em que o apartamento está localizado.	1 = 1º e 3º andar 2 = 4º e 6º andar 3 = 7º e 9º andar 4 = 10º andar ou mais	X	X	X
Conservação	Identifica o nível de conservação do imóvel.	1 = péssima 2 = regular 3 = boa 4 = ótima	X	X	X
idreal	Escore referente à idade cronológica do prédio (reflete o estado tecnológico).	1 = mais de 20 anos 2 = de 15 a 20 anos 3 = de 10 a 15 anos 4 = de 5 a 10 anos 5 = de 1 a 5 anos 6 = até 1 ano	X	X	X
idaparente	Escore referente à idade aparente do prédio.	(idem)	X	X	
descola	Identifica a proximidade com escolas.	1 = próximo até 500 metros 2 = de 500 a 800 metros 3 = mais de 800 metros	X	X	X
dhospital	Identifica a proximidade com hospitais.	(idem)	X	X	X
dsmercado	Identifica a proximidade com supermercados.	(idem)	X	X	X

Apêndice 1: Continuação...

Atributos	Descrição	Categorias	Aptos. Grupo A	Aptos. Grupo B	Aptos. Grupo C
Local	Classifica o bairro e outras características onde está localizada a residência.	1 = em valorização 0 = indiferente - 1 = em desvalorização			X
Posição do apartamento	Identifica a posição do apartamento em relação ao prédio (frente, lateral e fundos).	1 = frente 2 = lateral 3 = fundos	X	X	X
Acabamento	Identifica os vários níveis de acabamento.	1 = baixo 2 = normal 3 = alto	X	X	X
Pavimento	Número de apartamentos do prédio.	Quantidade	X	X	X
Elevador	Indica o número de elevadores do prédio.	Quantidade	X	X	X
Área const.	Indica a área do apartamento expressa em metros quadrados.	Metragem	X	X	X
Vgaragem	Indica o número de vagas para carros disponíveis para o apartamento.	Quantidade	X	X	
Quarto	Indica o número de dormitórios do apartamento.	Quantidade	X	X	X
Depempreg	Indica a existência (ou não) de dependência de empregados.	0 = inexistente 1 = existente		X	X
Suíte	Indica a presença (ou não) de suíte.	0 = inexistente 1 = existente			X
Sala	Indica o número de salas existentes no apartamento.	Quantidade	X	X	X
Peças	Indica o número de peças constituintes do apartamento.	Quantidade	X	X	X
Banheiro	Indica o número de banheiros do apartamento.	Quantidade	X	X	X
Total de atributos			17	19	19

Sobre os autores**Maria Teresinha Arns Steiner****Anselmo Chaves Neto****Sílvia Neide Braulio****Valdir Alves**

Universidade Federal do Paraná, Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Departamentos de Estatística e Matemática, CP 19081, CEP 81531-990, Curitiba, PR, Brasil, e-mails: anselmo@ufpr.br; tere@ufpr.br

Recebido em 27/11/2006

Aceito em 19/10/2007