TESTES ESTATÍSTICOS UTILIZADOS PARA A VALIDAÇÃO DE REGRESSÕES MÚLTIPLAS APLICADAS NA AVALIAÇÃO DE IMÓVEIS URBANOS

Statistical Tests Used in Multiple Regression Applied to the Urban Area Evaluation

CARLOS AURÉLIO NADAL [1]
KATIA APARECIDA JULIANO[2]
EDUARDO RATTON[3]

[1]UFPR – Departamento de Geomática
Curso de Pós-Graduação em Ciências Geodésicas
e-mail: cnadal@ufpr.br

[2]UFPR – Curso de Pós-Graduação em Ciências dos Solos
Instituto de Criminalística do Estado do Paraná.
e-mail: kjuliano@pop.com.br

[3]UFPR – Departamento de Transportes
e-mail: eduardo@seplan.brte.com.br

RESUMO

As regressões múltiplas são utilizadas para a determinação do valor de mercado de um imóvel urbano. As principais variáveis formadoras do valor são: a frente do lote, a localização, a área, o padrão construtivo das edificações, o estado geral de conservação e outras. Para se poder aplicar a regressão múltipla deve-se transformar uma variável não numérica como por exemplo, o tipo de padrão construtivo (baixo, médio e alto), em valor numérico, através a adoção de uma variável que o represente. O valor de mercado de um imóvel é então obtido por inferência estatística a partir do conhecimento dos valores de compra, venda e alugueis ou outra forma de geração de renda de outros imóveis que apresentem similaridades com aquele estudado. Neste trabalho será mostrado o resultado do tratamento do problema pelo método dos mínimos quadrados na forma matricial, enfatizando-se a detecção de erros grosseiros pelo método de Barda. Os critérios de aceitação e rejeição de hipóteses são aqueles prescritos em normas técnicas específicas. Um estudo de caso para um imóvel localizado no centro da Cidade de Curitiba é mostrado em detalhes neste trabalho.

ABSTRACT

This paper aims at presenting a solution to the problem of urban area evaluation. The mathematical model used in the problem is a multiple regression fitting method, that is tested in a study of real estate value in Curitiba, where the prices are obtained by comparing values from other properties for sale in the open market. The choice of the kind of variables to be used is very complex, but the expert's experience is very important to solve this kind of problem. In this paper the least square method with matrix models is applied, that is not usual in this kind of work in Brazil. The use of statistical hypothesis tests is very important in order to detect the reliability method. The blunders are detected by the Barda Method, by using residuals of estimated parameters. The confidence intervals obtained are the same of Brazilian standartization.

1. INTRODUÇÃO

A avaliação de imóveis é utilizada na grande maioria dos negócios, discussões e pendências interpessoais e sociais em nossas comunidades, tais como na compra ou na venda de casas, lojas comerciais, instalações industriais, aluguéis, na reavaliação de ativos de empresas, em atendimento à legislação vigente, na partilha oriunda de heranças, meações ou divórcios, no lançamento de impostos, nas hipotecas imobiliárias, nas divergências que originam ações demarcatórias, possessórias, nas indenizações, nas desapropriações e servidões, enfim, em um número expressivo de ações oriundas de problemas inerentes aos relacionamentos humanos, onde o valor de um bem assume importância fundamental.

Apesar do conceito de valor ser de difícil definição, sujeito e suscetível às mudanças filosóficas, torna-se importante no relacionamento humano e social adotar-se alguns critérios para que se exerça um caráter de justiça em sua aplicação prática. Assim, um trabalho de avaliação imobiliária constitui-se de uma seqüência de operações que resultam no que poderia ser chamado de uma "formação de juízo" sobre o valor de um imóvel ou um direito sobre ele.

A norma brasileira NBR5676 (ABNT, 1989) trabalha com o conceito de que o valor é aquele fornecido para um dado instante, único, não importando qual a finalidade da avaliação. Esse valor corresponde ao preço que se definiria, para um determinado imóvel, em um mercado de concorrência perfeito, sujeito às seguintes premissas:

a) homogeneidade dos bens levados a mercado;

b) números elevados de compradores e vendedores (o mercado não pode por eles ser alterado);

c) sem influência externa;

d) conhecimento pleno e absolutos sobre o mercado, sobre os bens e das tendências de avaliação por parte dos compradores e vendedores;

e) vendedores e compradores oferecendo liquidez com liberdade plena de entrada e saída do mercado.

A partir destas considerações, pode-se afirmar que a avaliação passa a ser a determinação técnica do valor ou de um direito sobre o imóvel. Dentre outros fatores deve-se levar em conta que o valor de um bem está diretamente ligado à sua capacidade de produzir renda, sua utilização potencial, o atendimento de uma necessidade ou a sua raridade.

Os imóveis urbanos podem ser definidos como bens que não são móveis, localizados nas cidades, geralmente classificados como glebas urbanizáveis, áreas ou lotes e terrenos com benfeitorias (casas, prédios residenciais, prédios comerciais, galpões e outros). A NBR 5676- ABNT (1989) define como sendo uma gleba urbanizável, "uma grande extensão de terreno passível de receber obras e infra-estruturas urbanas, por sua localização, seus aspectos físicos, sua destinação legal e pela existência de um mercado comprador".

Neste trabalho optar-se-á pelo método comparativo de dados de mercado, que é aquele que define o valor através da comparação com dados de mercado assemelhados, quanto às características intrínsecas e extrínsecas. Assim as características e os atributos dos dados pesquisados, que exerçam influência na formação do valor, devem ser homogeneizados por inferência estatística, respeitados os níveis de rigor definidos por norma técnica.

O nível de rigor mede a precisão do trabalho e será tanto maior quanto menor for a subjetividade contida na avaliação. O rigor de uma avaliação está condicionado à pesquisa efetivada, à confiabilidade dos dados coletados e à qualidade do modelo aplicado no processo de avaliação. A norma NBR 5676 classifica o trabalho nos seguintes níveis: expedito, normal, rigoroso ou rigoroso especial.

Torna-se muito importante a pesquisa dos dados utilizados neste método, com caracterizações objetivas e que sejam oriundos de regiões com as mesmas características sócio-econômicas, fornecidos por fontes seguras, com as respectivas épocas de oferta.

As variáveis utilizadas na inferência estatística merecem destaque e, para cada tipo de problema devem ser classificadas, estudadas e aceitas através de testes estatísticos. Assim, por exemplo, para se avaliar um lote urbano deve-se levar em conta algumas variáveis tais como: a dimensão de testada, a profundidade, a área total, a localização, o uso do solo, as posturas municipais, o zoneamento urbano, as distâncias a pólos que os valorizem ou os desvalorizem, a taxa de ocupação, a topografia, a suscetibilidade a enchentes ou a danos ambientais, o padrão de construções na vizinhança, a infra-estrutura urbana, a paisagem visual a partir do imóvel. Estas e outras variáveis permitem ao final a determinação do valor unitário do terreno pesquisado com relação à sua área total.

Estas variáveis devem ser ponderadas para que conceitos estatísticos possam ser aplicados com vistas a determinação do melhor modelo de ajuste.

Embora muitas publicações tratem deste assunto, poucas dão um tratamento matricial ao problema, sendo importante salientar que no tópico referente à detecção dos erros grosseiros, trabalham somente com critérios práticos, não aplicando testes estatísticos. Assim, visando dar um tratamento mais adequado a este tipo de erro está sendo proposta, no presente trabalho, a aplicação do método de Barda (“data snooping”).

2. REGRESSÕES MÚLTIPLAS

A regressão pode ser definida como sendo o estabelecimento de uma relação funcional entre duas ou mais variáveis envolvidas para a descrição de um fenômeno. A variável Y é aleatória e pode ser descrita matematicamente pela expressão [Marques, 2000]:

$$Y = f(X) + \varepsilon \qquad (1)$$

onde:

X é a variável independente ou variável explicativa;
Y é a variável dependente ou variável resposta;
ε é a componente aleatória da variação de Y;
f é a função de regressão.

Normalmente, X é uma variável que pode ser **controlada** pelo pesquisador, enquanto Y não é passível de controle. A análise de um gráfico de dispersão pode sugerir uma relação funcional entre as variáveis, como por exemplo, uma reta, uma exponencial, etc. Surge neste caso o modelo estatístico denominado de regressão linear simples. Uma generalização dessa regressão é conhecida como regressão múltipla. O modelo estatístico utilizado neste caso será dado por:

$$y_i = a + b_1 x_1 + b_2 x_2 + ... + b_u x_u + v_i \qquad (2)$$

onde, a, b_1, b_2 ... b_u são denominados de parâmetros da regressão múltipla, v substitui ε, e será denominado neste trabalho de resíduo. Cabe aqui alguma consideração estatística sobre este modelo, o qual pressupõe que a variável y_i é aleatória, que a esperança matemática dos resíduos é nula, ou seja, que a média dos resíduos é nula, que a variância de v_i é constante e igual a σ^2 (condição de homocedasticidade dos resíduos), que os erros são independentes entre si e que os mesmos tenham distribuição normal. Estatisticamente também se supõe que v_i é a componente aleatória da variação de Y, no entanto, devido a problemas práticos, alguns resíduos de erros sistemáticos e erros grosseiros, que interferem no processo, fazem com que v_i contenha também parte destes. Por este motivo, pode-se afirmar que o resíduo (v) compõe-se de três componentes: uma aleatória, uma sistemática e uma grosseira.

No problema proposto neste trabalho, y representa as observações (medidas) dos valores de imóveis em moeda corrente, x_i representa as variáveis, tais como: frente do lote, área, etc, formando um sistema de n equações denominado de equações de observações, cujas u incógnitas a, b_1, b_2, ... b_u, são objeto de determinação. O problema geralmente será resolvido para um número de observações maior que o número de incógnitas. De forma matricial, o sistema de equações de observação pode ser expresso como [Gemael, 1994]:

$$ {}_nA_u \, {}_uX_1 - {}_nL_1 = {}_nV_1 \qquad (3) $$

onde, ${}_nA_u$ é a matriz dos coeficientes das incógnitas, definida utilizando-se as derivadas parciais da função em relação às incógnitas e que pode ser escrita como:

$$ {}_nA_u = \begin{pmatrix} \partial f/\partial a & \partial f/\partial b_1 & \partial f/\partial b_2 & \dots & \partial f/\partial b_u \\ 1 & x_1 & x_2 & & x_u \\ 1 & x_1 & x_2 & & x_u \\ & \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ 1 & x_1 & x_2 & & x_u \end{pmatrix} $$

X_1 o vetor das incógnitas é dado por:

$$ {}_1X_u^T = \begin{bmatrix} a & b_1 & b_2 & \dots & b_u \end{bmatrix} $$

o vetor ${}_uL_1$ das variáveis respostas (observações) é dado por:

$$ {}_1L_n^T = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} $$

o vetor ${}_nV_1$ é o vetor dos resíduos estimados obtido pela expressão:

$$ {}_nV_1 = {}_nA_u \, {}_uX_1 - {}_nL_1 $$

O estimador dos parâmetros, por mínimos quadrados, será dado pela expressão:

$$ {}_1X_u = ({}_uA_n^T \, {}_nP_n \, {}_nA_u)^{-1} \, {}_uA_n^T \, {}_nP_n \, {}_nL_1 , $$

ou, reduzindo a simbologia,

$$ {}_1X_u = {}_uN_n^{-1} \, {}_uU_1 \qquad (4) $$

sendo P a matriz dos pesos, que na maioria dos casos de avaliações coincide com a matriz identidade (I), pois considera-se que as observações são provenientes de uma mesma população. No entanto, de forma generalizada tem-se:

$$ {}_nP_n = \sigma_o^2 \, {}_n\Sigma_{Lb\,n}^{-1} $$

sendo, σ_o^2 a variância da unidade de peso *a priori* para a qual arbitra-se o valor da unidade. O valor da variância da unidade de peso *a posteriori*, pode ser calculado pela expressão,

$$\sigma *_o^2 = \frac{{}_1V_n^T \; {}_nP_n \; {}_nV_1}{n - u} \tag{5}$$

A matriz variância-covariância dos parâmetros ajustados será dada por:

$${}_n\Sigma X^a{}_n = \sigma *_o^2 \; {}_uN_u^{-1} \tag{6}$$

3. TESTES ESTATÍSTICOS COMO DIRETRIZES EM ANÁLISE DE AJUSTAMENTO DE REGRESSÕES MÚLTIPLAS

Ao se analisar o ajustamento onde aplicam-se regressões múltiplas às avaliações, utilizam-se uma série de testes estatísticos com os mais variados objetivos, principalmente no que concerne à validação do modelo adotado. Esta questão é estudada através de testes de hipóteses, sendo que a NBR 5676 estabelece dois níveis de significância (α): nível rigoroso (5%) e nível rigoroso especial (1%); os quais são utilizados em todos os trabalhos de avaliação. Os principais testes efetivados são os seguintes:

a) Teste de qui-quadrado na verificação da bondade do ajustamento

É usado na verificação do ajustamento, utilizando como parâmetro para este tipo de aplicação um nível de significância (α) que tem como hipótese básica: (H_0) $\sigma_o^2 = \sigma *_o^2$, ou seja, a variância de unidade de peso, *a priori*, é igual estatisticamente à variância da unidade de peso *a posteriori*, no nível de significância que foi pré-definido, adotando-se ainda, como hipótese alternativa (H_1) $\sigma_o^2 \neq \sigma *_o^2$. O valor de qui-quadrado calculado (χ^2_c) é dado pela expressão:

$$\chi^2_c = \frac{\sigma *_o^2 \,(n\text{-}u)}{\sigma_o^2} \tag{7}$$

A distribuição qui-quadrado χ^2, em função dos graus de liberdade (n-u) e do nível de significância (α) fornece os valores referentes a:

$$\chi^2_{n\text{-}u,\,0,5\alpha} \text{ e } \chi^2_{n\text{-}u,\,1\text{-}0,5\alpha}$$

A hipótese H_0 é aceitável se:

$$\chi^2_{n\text{-}u,\,0,5\alpha} < \chi^2_c < \chi^2_{n\text{-}u,\,1\text{-}0,5\alpha} \tag{8}$$

b) Coeficiente de correlação linear múltiplo (R) e coeficiente de determinação (R^2)

O coeficiente de correlação traduz numericamente o quanto as variáveis estão linearmente relacionadas entre si. É fornecido matricialmente pela raiz quadrada da expressão:

$$R^2 = \frac{{}_1X_u^T\, {}_uA_n^T\, {}_nP_n\, {}_nZ_1}{{}_1Z_n^T\, {}_nZ_1} \tag{9}$$

com

$${}_nZ_1 = {}_nL_1 - {}_nL^*_1$$

todos os elementos do vetor ${}_nL^*_1$ são iguais e obtidos pela média aritmética dos n valores de y.

O valor de R encontra-se no intervalo:

$$-1 \leq R \leq 1$$

já o coeficiente de determinação indica numericamente o percentual do valor de avaliações que esta sendo explicitado pelo modelo, encontra-se no intervalo:

$$0 \leq R^2 \leq 1$$

c) Teste de existência da regressão:

Consiste em se estudar a probabilidade dos parâmetros de regressão a, b_1, b_2... b_u serem iguais a zero ao mesmo tempo, neste caso não existe regressão. O teste é efetivado através da distribuição Fischer-Snedecor. O coeficiente F calculado é obtido pela expressão matricial:

$$F_c = \frac{({}_1X_u^T\, {}_uA_n^T\, {}_nP_n\, {}_nZ_1)(n\text{-}k\text{-}1)}{(\,{}_1Z_n^T\, {}_nZ_1 - {}_1X_u^T\, {}_uA_n^T\, {}_nP_n\, {}_nZ_1)\, k} \tag{10}$$

com

$$k=u-1$$

A hipótese básica é aceita, ou seja, de que haja regressão de y em x_1, x_2, etc. com nível de significância (α) se:

$$F_c > F$$

d) Teste da significância dos regressores

Utiliza-se neste caso a distribuição T de Student (unicaudal), tendo como hipótese básica que os regressores são diferente de zero a um nível de significância α. Os valores calculados para o modelo de melhor ajuste são dados pela expressão:

$$T_{bi} = \frac{b_i}{\sigma_{bi}}$$

com base nos valores obtidos da distribuição de Student (T) com entradas 2α e graus de liberdade n-k-1, aceita-se a hipótese básica que cada um dos coeficientes b_i seja diferente de zero se:

$$T_{bi} > T$$

e) Teste "*data snooping*" - Barda

O teste de Barda pode ser definido como a investigação em relação à observação na qual um erro grosseiro foi cometido [Barda, 1968], está baseado no teste estatístico de resíduos padronizados após o ajustamento por mínimos quadrados [Moraes, 1997]. A detecção de erros grosseiros foi utilizada na Fotogrametria por Mitishita [1980], que classificou os tipos de erros grosseiros:

Tipo	Nome do erro	Magnitude
1	"blunders"	$m > 170\sigma$
2	"blunders"	$m \leq 170\sigma$
3	"outliers"	$3\sigma < m < 100\sigma$

Para que se possa aplicar o teste de Barda devem ser observadas as seguintes condições:

- deve ser aplicado após o ajustamento, portanto, os cálculos devem estar rigorosamente corretos;
- os pesos devem ser escolhidos apropriadamente para evitar-se a distribuição dos erros grosseiros nos resíduos;
- o nível de confiança adotado deve ser o mesmo adotado no teste de qui-quadrado para verificação da bondade do ajustamento.

Calcula-se a matriz de redundância ou dos coeficientes de peso dos resíduos dada pela expressão [Moraes, 1997]:

$$_{n}Q_{vn} = {}_{n}I_{n} - {}_{n}A_{u}\, {}_{u}N_{n}^{-1}\, {}_{u}A_{n}^{T}\, {}_{n}P_{n} \qquad (11)$$

O número de redundância é obtido da diagonal principal da matriz fornecida pelo resultado do produto [Moraes, 1997];

$$r_i = [{}_{n}Q_{vn}\, {}_{n}P_{n}]_{ii} \qquad (12)$$

Os resíduos padronizados serão dados pela expressão:

$$w_i = \frac{v_i \quad \sqrt{p_i}}{\sigma_0 \quad \sqrt{r_i}}$$

A hipótese básica é que não há nenhum erro grosseiro na observação e é rejeitada se:

$$w_i > K$$

onde, K é o valor crítico de acordo com um nível de confiança específico conforme a tabela [Moraes, 1997], onde a probabilidade α de erro tipo I (rejeição de Ho quando verdadeira) e a probabilidade β de erro tipo II (aceitação de Ho quando falsa) sejam tão pequenas quanto possível:

1-α	K	1-β
99,9%	3,29	76%
99,7%	3,00	84%
99,0%	2,56	93%
95,0%	1,96	98%

f) Teste de Durbin-Watson

Existe autocorrelação ou correlação serial quando os termos de resíduos são correlacionados com os valores anteriores ou posteriores da mesma série. A má especificação do modelo de regressão, em função de resíduos na forma do modelo ou por exclusão de variáveis independentes importantes para a análise é uma das causas da autocorrelação. Isto ocorre principalmente em aplicações envolvendo séries temporais [Johnston, 1977].

A autocorrelação pode ser verificada pela denominada estatística de Durbin-Watson [IMAPE, 1998], onde a hipótese básica é a existência de autocorrelação entre resíduos, que pode ser calculada pela expressão:

$$dw = \frac{\Sigma[v(i)-v(i-1)]^2}{{}_1V_n^T \; {}_nP_n \; {}_nV_1} \qquad (13)$$

Da estatística de Durbin-Watson para o nível de confiança α com v =n-k-1, obtém-se du que é o limite superior de variação e di, o limite inferior, assim: se dw> du a hipótese básica é aceita, se dw<di é rejeitada e se ocorrer que di<dw<du o teste é inconclusivo.

g) Teste de Kolmogorov-Smirnov (normalidade dos resíduos)

A condição de normalidade dos resíduos não é necessária para a obtenção dos estimadores pelo método dos mínimos quadrados, mas sim para a definição de intervalos de confiança e testes de significância. A falta de normalidade é uma indicação de que os estimadores são não tendenciosos.

Antes de se aplicar um teste, compara-se a distribuição dos resíduos obtidos com a distribuição normal. Para se efetuar esta comparação deve-se homogeneizar os resíduos calculados dividindo-os pelo desvio padrão obtido a *posteriori* (resultado positivo da raiz quadrada da variância a *posteriori*). A comparação das distribuições nos permite analisar se os resíduos obtidos apresentam ou não uma distribuição aproximada da curva normal, atestando ou não a hipótese de normalidade dos resíduos, que pode ser considerada como uma verificação ou inspeção visual subjetiva da normalidade.

O teste de Kolmogorov-Smirnov avalia se duas amostras tem distribuições semelhantes, ou melhor dizendo, se foram extraídas de uma mesma população. Se apresentarem grandes diferenças provavelmente estas não se devem ao acaso. É um teste que detecta diferenças em relação à tendência central, dispersão e simetria.

Para se aplicar o teste deve-se ordenar as amostras, construir as distribuições de freqüências acumuladas nos intervalos de classe definidos, calcular as diferenças entre estas freqüências (da primeira menos a da segunda amostra), escolhendo-se a maior diferença em valor absoluto ($d_{máx}$) que será comparada com um valor tabelado ($d_{crítico}$). A hipótese básica do teste é que se $d_{máx} \geq d_{crítico}$ rejeita-se a igualdade das amostras.

Para maiores detalhes sobre o teste de Kolmogorov-Smirnov sugere-se Bunchaft (1997).

h) Análise gráfica dos resíduos

Ao se colocar em um gráfico os resíduos e as variáveis explicativas é possível a verificação da existência de uma multi-colinearidade, ou seja, uma relação exata entre as variáveis, se o coeficiente de correlação apresenta-se muito próximo de um. Se o gráfico demonstrar que os resíduos não estão alinhados então a correlação é meramente casual e os resíduos não mostram nenhuma tendência.

Ao se analisar graficamente as distribuições (resíduos versus valor estimado) verifica-se a existência de homocedasticidade, ou seja, a hipótese de variância constante, que é aceita quando não há nenhuma tendência dos resíduos em relação ao valor estimado, neste caso denominado de hetereocedasticidade.

4. ESTUDO DE CASO

No ano de 1995, uma construtora de Curitiba, utilizando-se de legislação municipal referente a aquisição de potencial construtivo, aprovou um projeto para construção de um edifício residencial no centro da cidade, com cinco pavimentos, em uma área de zoneamento ZR-3, a qual permite a construção de, no máximo, dois pavimentos.

Com o desenvolvimento da obra, percebeu-se que a mesma influenciaria negativamente na paisagem urbana, principalmente porque inviabilizaria a visitação pública de turistas a um mirante de observação da Cidade.

O imóvel foi então desapropriado com base no Código de Posturas de Curitiba e seu valor, para fins de desapropriação, utilizando-se o método comparativo com dados de mercado [Lima, 2001].

O imóvel desapropriado possuía as seguintes características:

- terreno urbano com área total de 418,00 m^2, com forma irregular, de esquina, com testadas de 16,30m e 20,00m, lateral esquerda de 22,00m e fundos de 25,50m, com topografia regular e nível abaixo da rua;
- área construída total de 1032,32 m^2, de alvenaria, para fins residenciais com cinco pavimentos e um subsolo. Cada apartamento seria composto por 2 quartos e 1 suíte, sala com dois ambientes, banheiro social, banheiro de empregada e área de serviço. O apartamento possuiria duas garagens. O prédio teria dois elevadores. Foram executados cerca de 48% da obra quando desapropriada.

Utilizou-se a pesquisa amostral efetivada por Lima [2001] em diversas fontes tais como: instituições de classe, imobiliárias e jornais, considerando-se as características semelhantes ao imóvel estudado, tanto em termos de proximidade de localização, quanto de características próprias, entre elas: o zoneamento, a infra-estrutura urbana, o coeficiente de aproveitamento, o padrão construtivo, o número de quartos, etc.

Para os imóveis considerados como amostra, cada dado foi verificado até o grau de detalhamento que possibilita as condições de cotejá-lo com o bem imóvel avaliando, fixando-se o nível de rigor do trabalho com base na NBR-5676.

As amostras podem ser visualizadas no quadro a seguir [Lima, 2001]:

Nº Amostra	Valor unitário	Área equivalente	Padrão construtivo	Idade aparente	Vagas de garagens
1	857,46	590,00	Alto	0 a 1 ano	>2 vagas cobertas
2	733,00	400,00	Médio	4 a 6 anos	> 2 vagas cobertas
3	734,66	322,00	Médio	4 a 6 anos	2 vagas cobertas
4	737,22	223,00	Médio	8 a 10 anos	2 vagas cobertas
5	805,74	235,00	Alto	> 12 anos	2 vagas cobertas
6	762,47	152,03	Médio	6 a 8 anos	2 vagas cobertas
7	770,73	149,23	Médio	4 a 6 anos	2 vagas cobertas
8	735,71	140,00	Médio	10 a 12 anos	1 vaga coberta
9	634,45	163,00	Médio	> 12 anos	1 vaga coberta
10	630,85	155,00	Médio	> 12 anos	1 vaga coberta
11	472,17	180,02	Baixo	> 12 anos	2 vagas cobertas
12	464,57	132,15	Baixo	> 12 anos	1 vaga coberta
13	629,76	123,00	Médio	> 12 anos	1 vaga coberta
14	462,81	134,35	Baixo	> 12 anos	1 vaga coberta
15	648,04	89,50	Médio	> 12 anos	1 vaga coberta
16	513,11	107,19	Baixo	4 a 6 anos	1 vaga coberta
17	899,28	139,00	Alto	4 a 6 anos	1 vaga coberta
18	727,78	126,00	Médio	8 a 10 anos	1 vaga coberta
19	639,65	107,74	Médio	> 12 anos	1 vaga coberta
20	586,32	85,50	Baixo	0 a 1 ano	1 vaga coberta

I) Modelo de regressão adotado

Foi adotado o seguinte modelo de regressão linear múltipla

$$Y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + v_i$$

Onde as variáveis envolvidas são:

a) Variável Dependente:

Valor do Imóvel [Y]: valor unitário do imóvel, calculado por metro quadrado de área.

b) Variáveis Independentes:

Área equivalente[x_1]: = definida como a área equivalente de construção, que é calculada multiplicando-se a área dos componentes do imóvel pelos fatores mostrados no quadro abaixo:

Descrição da área	**Fator de área equivalente**
Pavimento tipo	1,00
Térreo coberto fechado	1,00
Térreo coberto aberto (pilotis)	0,50 a 0,70
Subsolo	0,50
Atico	0,75
Mezanino	0,90
Sobre Loja	0,80
Áreas complementares	0,15
Terraço coberto	0,50
Piscinas	1,00
Vagas de Garagem	1,00
Cobertas no Térreo	0,50
Descobertas sobre a Laje	0,30
Descobertas sobre a Terra	0,15

Inverso do Fator de Padrão Construtivo [x_2]: O fator de padrão construtivo foi obtido pela linearização da variação do valor da construção e acabamento entre o imóvel avaliando e os elementos que compõe a amostra, resultando para padrão baixo=408,23; padrão médio=483,93; alto padrão=585,51. Utilizou-se o inverso destes valores por experiência prática usual na avaliação de imóveis urbanos.

Inverso da Idade Aparente do imóvel [x_3]: classificaram-se as amostras com fatores de idade aparente: mais que (>) 12 anos = 1; de 10 a 12 anos = 3; de 8 a 10anos = 3; de 6 a 8 anos = 4; de 4 a 6 anos = 5; de 2 a 4 anos = 6; de 1 a 2 anos = 7; de 0 a 1 ano = 8.

Inverso do fator de vagas de garagem [x_4]: efetuou-se a seguinte classificação: garagem descoberta = 1; vaga coberta = 2; duas vagas cobertas = 3; mais que (>) duas vagas cobertas = 4).

II) Ajustamento pelo método dos mínimos quadrados

Para a solução pelo método dos mínimos quadrados é necessário a montagem numérica das matriz dos coeficientes das incógnitas (A). A matriz dos pesos (P) no problema em pauta foi igualada a matriz identidade, não será representada de forma explícita. O vetor das variáveis explicitadas (L) foi montado com os valores observados.

A matriz A e o vetor L são mostrados a seguir.

$$A = \begin{pmatrix}
1 & 590,00 & 0,00170791275981623 & 0,125 & 0,25 \\
1 & 400,00 & 0,00206641456408985 & 0,2 & 0,25 \\
1 & 322,00 & 0,00206641456408985 & 0,2 & 0,33333333 \\
1 & 223,00 & 0,00206641456408985 & 0.3333333 & 0,33333333 \\
1 & 235,00 & 0,00170791275981623 & 1 & 0,33333333 \\
1 & 152,03 & 0,00206641456408985 & 0,25 & 0,33333333 \\
1 & 149,23 & 0,00206641456408985 & 0,2 & 0,33333333 \\
1 & 140,00 & 0,00206641456408985 & 0,3333333 & 0,5 \\
1 & 163,00 & 0,00206641456408985 & 1 & 0,5 \\
1 & 155,00 & 0,00206641456408985 & 1 & 0,5 \\
1 & 180,02 & 0,00244959949048331 & 1 & 0,33333333 \\
1 & 132,15 & 0,00244959949048331 & 1 & 0,5 \\
1 & 123,00 & 0,00206641456408985 & 1 & 0,5 \\
1 & 134,35 & 0,00244959949048331 & 1 & 0,5 \\
1 & 89,50 & 0,00206641456408985 & 1 & 0,5 \\
1 & 107,19 & 0,00244959949048331 & 0,2 & 0,5 \\
1 & 139,00 & 0,00170791275981623 & 0,2 & 0,5 \\
1 & 126,00 & 0,00206641456408985 & 0,3333333 & 0,5 \\
1 & 107,74 & 0,00206641456408985 & 1 & 0,5 \\
1 & 85,50 & 0,00244959949048331 & 0,125 & 0,5
\end{pmatrix}$$

$$L=\begin{pmatrix} 857,46 \\ 733,00 \\ 734,66 \\ 737,22 \\ 805,74 \\ 762,47 \\ 770,73 \\ 735,71 \\ 634,45 \\ 630,85 \\ 472,17 \\ 464,57 \\ 629,76 \\ 462,81 \\ 648,04 \\ 573,11 \\ 899,28 \\ 727,78 \\ 639,65 \\ 586,32 \end{pmatrix}$$

Pela solução matricial da expressão (4), obtém-se o vetor (X) de estimativa dos parâmetros solicitados no modelo de regressão, que resultará em:

$$X=\begin{pmatrix} 1824,29715989785 \\ -0,214930982557917 \\ -454764,932555851 \\ -136,132168924983 \\ -168,353398806299 \end{pmatrix}$$

ou seja, o modelo de regressão resultante para o problema será dado pela expressão:

$$y = 1824,29715989785 - 0,214930982557917x_1 - 454764,932555851x_2 - 136,132168924983\,x_3 - 168,353398806299\,x_4$$

O vetor dos resíduos (V), resultará em:

$$V = \begin{pmatrix} 4,22417834238195 \\ -3,72289648263074 \\ -2,64772974363814 \\ -2,08051832706883 \\ -0,90042055944263 \\ -0,73251908451823 \\ -1,58410388710683 \\ -10,7901465758116 \\ -5,22833845796572 \\ 0,09110940249763 \\ 7,1933688122088 \\ -2,97678485379362 \\ 8,058900844351 \\ -1,68963301542095 \\ -3,02091123995865 \\ 2,75362761083841 \\ 7,03978910505396 \\ 0,148887179999292 \\ 1,44874763818473 \\ 4,4153932918933 \end{pmatrix}$$

A matriz variância-covariância resultará em:

$$\Sigma X= \begin{pmatrix} 297,684711 & - 0,224692 & -75736,110107 & -0,997704 & -220,937116 \\ -0,224692 & 0,000278 & 32,166531 & 150,004464 & 0,2400921 \\ -75736,110107 & 32,166531 & 32013965,173475 & -371,031620 & 5677,3653844 \\ -0,997704 & 0,0044645 & -371,031620 & 10,6794342 & -12,231989 \\ -220,937116 & 0,240092 & 5677,365384 & -12,231989 & 402,207891 \end{pmatrix}$$

III) Testes estatísticos

Uma análise do ajustamento é efetivada através de testes estatísticos, conforme proposto no trabalho:

a) Teste de qui-quadrado para verificação da bondade do ajustamento.

O valor da variância da unidade de peso a posteriori resultou a partir da expressão (5) em:

$$\sigma{*}_o^2 = 27,1307176459636$$

O valor de qui-quadrado calculado pela (7) resulta em:

$\chi^2_c = 406{,}960764689454$

Na distribuição qui-quadrado para um nível de significância de 95% tem-se:

$\chi^2_{15,\,0,025} = 6{,}26$

$\chi^2_{15,\,9,725} = 27{,}49$

Verifica-se neste caso que a hipótese básica deve ser rejeitada, deve-se proceder a uma análise aprofundada do ajustamento (pode-se considerar os resíduos excessivamente grandes em decorrência de erros grosseiros ou sistemáticos). Deve-se também analisar se o modelo matemático utilizado é consistente com as observações. O ideal seria parar o ajustamento neste ponto e efetuar um estudo sobre diferentes modelos. Por se tratar de um estudo de caso prossegue-se com o tratamento dos erros.

b) coeficiente de correlação linear múltiplo (R) e coeficiente de determinação (R^2)

Calculou-se também o coeficiente de determinação que resultou em:

$R^2 = 0{,}998601862578729$

cujo significado é que 99,86% do valor de mercado está sendo explicitado pelo modelo e, o coeficiente de correlação linear múltiplo resulta em:

$R = 0{,}999300686769868$

que mostra uma correlação fortíssima entre as variáveis explicativas e a explicada.

c) Teste de existência da regressão

O coeficiente F calculado é obtido pela expressão matricial (10) resultando em:

$Fc=2678{,}3897832203$

O valor tabelado para o nível de significância de 5% resultou em:

$F = 3{,}056$

como $F_c > F$ aceita-se a hipótese de existência de regressão.

d) Teste da significância dos regressores

Nesse caso deve-se dividir cada um dos regressores pela raiz quadrada do elemento da diagonal principal da matriz variância covariância, que representa o erro médio quadrático deste, assim obtém-se:

$T_{b1} = -12{,}8797699083259$

$T_{b2} = -80{,}3743056845373$

$T_{b3} = -41{,}6568706036683$

$T_{b4} = -8{,}39453405047628$

O valor do coeficiente T de Student retirado da tabela da distribuição com um nível de significância de 5% resultou em:

$T = 2{,}1314$

Como os valores calculados em módulo são superiores ao valor critico aceita-se a hipótese básica de significância dos regressores.

e) Teste "*data snooping*" - Barda

Calcula-se a matriz de redundância ou dos coeficientes de peso dos resíduos utilizando-se a expressão matricial (11), que resulta em:

0,3000 -0,3026 -0,2345 -0,0298 -0,0458 0,1122 0,1165 –0,0420 -0,0710 -0,0548 0,0852
0,0017 0,0102 -0,0028 0,0783 0,0311 -0,0529 -0,0136 0,0412 0,0732
-0,3026 0,7374 -0,1759 -0,1201 -0,0430 -0,0860 -0,0862 0,0155 0,0257 0,0299 -0,1410
-0,0236 0,0465 -0,0248 0,0639 -0,0372 0.,728 0,0228 0,0544 -0,0285
-0,2345 -0,1759 0,8605 -0,0832 0,0013 -0,0578 -0,0607 –0,0437 0,0016 0,0052 -0,0445
-0,0210 0,0198 -0,0220 0,0351 -0,0765 -0,0201 -0,0373 0,0268 -0,0729
-0,0298 -0,1201 -0,0832 0,8698 -0,1006 -0,1723 -0,1762 -0,0171 0,0268 0,0225 -0,1127
0,0193 0,0053 0,0205 –0,0127 -0,0321 -0,0325 -0,0247 -0,0029 -0,0473
-0,0458 -0,0430 0,0013 -0,1006 0,6020 -0,1339 -0,1249 0,0414 -0,0897 -0,0956 -0,1107
0,0569 -0,1189 0,0585 –0,1434 0,2162 -0,0880 0,0312 -0,1301 0,2170
0,1122 -0,0860 -0,0578 -0,1723 -0,1339 0,7291 -0,2790 –0,0225 0,0682 0,0580 -0,1253
0,0696 0,0170 0,0724 -0,0259 -0,0359 -0,0742 -0,0404 -0,0025 -0,0706
0,1165 -0,0862 -0,0607 -0,1762 -0,1249 -0,2790 0,7117 -0,0291 0,0759 0,0654 -0,1163
0,0771 0,0232 0,0800 –0,0209 -0,0454 -0,0845 -0,0475 0,0031 -0,0824
-0,0420 0,0155 -0,0437 -0,0171 0,0414 -0,0225 -0,0291 0,8582 -0,0529 -0,0523 0,1107
-0,0169 -0,0495 -0,0171 –0,0467 -0,1238 -0,1911 -0,1406 -0,0483 -0,1322
-0,0710 0,0257 0,0016 0,0268 -0,0897 0,0682 0,0759 -0,0529 0,8505 -0,1460 -0,0149
-0,1098 -0,1323 -0,1107 -0,1179 0,0050 -0,0599 -0,0469 -0,1257 0,0240
-0,0548 0,0299 0,0052 0,0225 -0,0956 0,0580 0,0654 -0,0523 -0,1460 0,8567 -0,0182
-0,1052 -0,1321 -0,1060 -0,1204 0,0064 -0,0629 -0,0474 -0,1268 0,0236
0,0852 -0,1410 -0,0445 -0,1127 -0,1107 -0,1253 -0,1163 0,1107 -0,0149 -0,0182 0,5931
-0,1689 -0,0315 -0,1680 -0,0454 -0,0171 0,2694 0,1049 -0,0378 -0,0109
0,0017 -0,0236 -0,0210 0,0193 0,0569 0,0696 0,0771 -0,0169 -0,1098 -0,1052 -0,1689
0,7749 -0,0871 -0,2263 -0,0680 -0,1152 0,1238 -0,0090 -0,0784 -0,0939
0,0102 0,0465 0,0198 0,0053 -0,1189 0,0170 0,0232 -0,0495 -0,1323 -0,1321 -0,0315
-0,0871 0,8685 -0,0871 –0,1308 0,0122 -0,0748 -0,0493 -0,1312 0,0218
-0,0028 -0,0248 -0,0220 0,0205 0,0585 0,0724 0,0800 -0,0171 -0,1107 -0,1060 -0,1680
-0,2263 -0,0871 0,7724 -0,0673 -0,1156 0,1247 -0,0089 -0,0781 -0,0937
0,0783 0,0639 0,0351 -0,0127 -0,1434 -0,0259 -0,0209 -0,0467 -0,1179 -0,1204 -0,0454
-0,0680 -0,1308 -0,0673 0,8583 0,0182 -0,0873 -0,0512 -0,1358 0,0200
0,0311 -0,0372 -0,0765 -0,0321 0,2162 -0,0359 -0,0454 –0,1238 0,0050 0,0064 -0,0171
-0,1152 0,0122 -0,1156 0,0182 0,7299 -0,0326 -0,1213 0,0149 -0,2812
-0,0529 0,0728 -0,0201 -0,0325 -0,0880 -0,0742 -0,0845 –0,1911 -0,0599 -0,0629 0,2694
0,1238 -0,0748 0,1247 -0,0873 -0,0326 0,6013 -0,1963 -0,0805 -0,0545
-0,0136 0,0228 -0,0373 -0,0247 0,0312 -0,0404 -0,0475 -0,1406 -0,0469 -0,0474 0,1049
-0,0090 -0,0493 -0,0089 -0,0512 -0,1213 -0,1963 0,8586 -0,0502 -0,1329
0,0412 0,0544 0,0268 -0,0029 -0,1301 -0,0025 0,0031 -0,0483 -0,1257 -0,1268 -0,0378
-0,0784 -0,1312 -0,0781 -0,1358 0,0149 -0,0805 -0,0502 0,8667 0,0210
0,0732 -0,0285 -0,0729 -0,0473 0,2170 -0,0706 -0,0824 -0,1322 0,0240 0,0236 -0,0109
-0,0939 0,0218 -0,0937 0.0200 -0,2812 -0,0545 -0,1329 0,0210 0,7002

já, o número de redundância obtido é: r = 15,00, resultando para os resíduos padronizados os seguintes valores:

$$
W = \begin{pmatrix} 0,2094 \\ -0,1845 \\ -0,1312 \\ -0,1031 \\ -0,0446 \\ -0,0363 \\ -0,0785 \\ -0,5349 \\ -0,2592 \\ 0,0045 \\ 0,3566 \\ -0,1476 \\ 0,3995 \\ -0,0838 \\ -0,1497 \\ 0,1365 \\ 0,3490 \\ 0,0074 \\ 0,0718 \\ 0,2189 \end{pmatrix}
$$

Para um nível de significância de 5% retira-se k=1,96 (distribuição F de Snedecor). Como não há resíduos padronizados maior que k, aceita-se a hipótese básica de que não há erros grosseiros nas observações.

f) Teste de Durbin-Watson

Obtém-se pela expressão (13) o seguinte valor para essa estatística:

$$dw = 1,70266906699683$$

O limite inferior obtido das tabelas foi

$$di = 0,90$$

Já,. o limite superior

$$du = 1,83$$

Neste caso o teste mostra-se inconclusivo. Deve-se salientar que a autocorrelação só pode ser verificada se as amostragens estiverem ordenadas segundo um critério específico.

g) Teste de normalidade dos resíduos

Uma verificação qualitativa da normalidade pode ser analisada com a distribuição dos resíduos padronizados em relação ao erro médio quadrático. Neste caso nota-se que 80% dos resíduos distribuem-se no intervalo definido por mais ou menos uma vez o erro médio quadrático e 95% no intervalo de ± 1,96 vezes o erro que corresponde a um intervalo de 95% na distribuição de Gauss.

A aplicação do teste de normalidade de Kolmogorov-Smirnov foi efetivada utilizando-se um nível de significância de 95%, obtendo-se que a distribuição dos resíduos é normal.

No presente caso, analisaram-se os gráficos de resíduos das observações. cujas distribuições apresentam-se aleatórias, mostrando que há homocedasticidade e não ocorre multi-colinearidade.

5. CONCLUSÕES E RECOMENDAÇÕES

Considera-se uma contribuição a este tipo de problema a utilização de uma solução matricial pela aplicação do método dos mínimos quadrados, o que não é usual nos trabalhos de avaliações consultados pelos autores. A utilização do método de Barda para analise da existência de erros grosseiros é um dos fundamentos teóricos apresentados para o problema neste trabalho. O uso da matriz variância-covariância permite que os cálculos envolvidos sejam simplificados. Outras análises podem ser conduzidas, através de outros testes estatísticos que melhoram a compreensão do problema.

A metodologia em análise é de fácil aplicação. A confiabilidade de resultados em uma avaliação é fator preponderante no trabalho, uma vez que este resultado será utilizado tanto na solução de conflitos, como na determinação de coeficientes de zoneamento fiscal para distribuição de tributos. Por último, é fundamental que a metodologia apresentada na solução de problemas de avaliação mostre de forma definitiva a imparcialidade da avaliação e neste caso uma análise criteriosa de erros é fundamental para as validações dos trabalhos.

AGRADECIMENTOS

Os autores desejam expressar seus agradecimentos ao Prof. Dr Jair Mendes Marques pelas sugestões e correções ao texto.

REFERÊNCIAS

ABNT. **NBR 5676 - Avaliação de Imóveis urbanos**. Rio de Janeiro, 1989.

BAARDA, W. A. **A testing procedure for use in geodetic networks**. *Netherlands Geodetic Comission,* v. 2, n. 5, 1968.

BUNCHAFT, G. **Estatística sem mistérios.** 4ª ed. Petrópolis, RJ. Vozes, 1997.

GEMAEL, C. **Introdução ao ajustamento de observações: aplicações geodésicas**. 1. ed. Curitiba: Editora UFPR. 1994.

IMAPE. **Fundamentos de Avaliações Patrimoniais e Perícias de Engenharia.** 1. ed. São Paulo. Editora Pini. 1998.

JOHNSTON, J. **Métodos econométricos.** São Paulo: Atlas, 1977.

LIMA, A. J. M. **Determinação do valor de indenização para desapropriação de edifício residencial em construção, para fins de preservação do patrimônio paisagístico.** Seminário do Curso de Pós-Graduação em Construção Civil da Universidade Federal do Paraná, 2001.

MARQUES, J. M. **Estatística - Curso de Engenharia**. UFPR, Curitiba, 2000.

MITISHITA, E. A. **Detecção de erros grosseiros nas aerotriangulações.** Curitiba, 1986. Dissertação (Mestrado em Ciências Geodésicas) Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná.

MORAES, C. V. **Aplicação do ajustamento às poligonais.** Curitiba, 1997. Dissertação (Mestrado em Ciências Geodésicas) Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná.

(Recebido em setembro/03. Aceito em novembro/03)