

# Modelagem do Mercado de Terras por Análise de Regressão Múltipla para definição de Planta de Valores Genéricos

Dr. Everton da Silva <sup>1</sup>  
Dr. Carlos Loch <sup>2</sup>  
Dr<sup>a</sup>. Liane Silva Ramos <sup>3</sup>

<sup>1</sup> Tekoha Engenharia e Consultoria Ltda.  
Blumenau SC  
veto68@terra.com.br ou everton@tekoha.com.br

<sup>2</sup> Universidade Federal de Santa Catarina—UFSC.  
Caixa postal 476 Campus Universitário—Florianópolis SC  
loch@ecv.ufsc.br

<sup>3</sup> Consultora em Cadastro Técnico Multifinalitário.  
Blumenau SC  
liane.sr@terra.com.br

**Resumo:** neste trabalho são apresentados procedimentos para modelagem do mercado de terras visando a definição de valores unitários para formar uma planta de valores genéricos. As análises foram baseadas em dados existentes no cadastro técnico, de onde se originou um conjunto de variáveis, algumas na forma direta, outras com algumas transformações para se ter uma escala adequada e outras resultantes de análises espaciais. Destaca-se a importância de se ter uma base de dados do mercado imobiliário integrada ao cadastro técnico, contendo eventos dos mais variados tipos de imóveis e em constante inclusão de novos eventos. A modelagem foi realizada com a utilização de análise de regressão múltipla.

**Palavras chaves:** Cadastro Técnico Multifinalitário, Avaliação em Massa, Regressão Múltipla.

**Abstract:** in this work procedures for modeling of the land market are presented aiming at the definition of unitary values to form a plant of generic values. The analyses had been based on existing data in it register in technical cadastre, from where it originated a set of variable, some in the direct form, others with some transformations to have an adequate scale and other resultants of space analyses. Importance is distinguished it of if to have a database of the integrated real estate market to it I register in technical cadastre, contend events of the most varied types of property and in constant inclusion of new events. The modeling was carried through with the use of analysis of multiple regression.

**Key words:** Multi-purpose Technical Cadastre, Mass Appraisal, Multiple Regression

## 1. Considerações Iniciais

O presente trabalho é parte do resultado de uma pesquisa desenvolvida em uma área de estudo na cidade de Blumenau - SC, que teve como objetivo principal avaliar e demonstrar o potencial do cadastro técnico para avaliação em massa de imóveis.

Tomando em conta as variáveis que podem ser obtidas a partir do cadastro técnico, de forma direta ou derivadas de análises espaciais, foi desenvolvido um modelo de regressão múltipla para os terrenos, onde sua aplicação se demonstra eficiente para a definição dos valores unitários de uma planta de valores genéricos. Apresentam-se os procedimentos utilizados, desde a análise exploratória dos dados até a modelagem.

Na referida área de estudo foi realizada uma pesquisa no mercado imobiliário visando constituir uma base de eventos relacionada ao cadastro técnico. Os eventos foram obtidos em fontes ofertas e de transações (ITBI – Imposto sobre a Transmissão de Bens Imóveis), e totalizaram 1.538 unidades distribuídas em 1.431 lotes. Trata-se de eventos em terrenos vagos ou baldios.

## 2. Metodologia Empregada para Modelagem

### 2.1. Análise exploratória do conjunto de dados de mercado

A análise exploratória do conjunto de dados de mercado armazenado no banco de dados do mercado imobiliário foi realizada com os objetivos de avaliar as características relacionadas a cada evento de mercado e de retirar àqueles que poderiam prejudicar o processo de modelagem, como os eventos duplos e casos de valores tidos como atípicos.

Embora o mercado imobiliário esteja bem organizado no sentido de que um imóvel seja ofertado por uma única imobiliária ou corretor num dado momento, ocorre que um mesmo imóvel em momentos distintos é ofertado por fontes distintas e com valores semelhantes ou mesmo iguais. Em poucos casos se permitiu a permanência destes casos na amostra, ficando aqueles que tinham valores distintos e um período de tempo entre as ofertas considerável.

Outra situação que motivou a análise exploratória foi a diferença nas dimensões de um mesmo imóvel entre a fonte de informações do mercado e o cadastro técnico. É importante enfatizar que também se tratavam de poucos casos, pois os dados cadastrais estavam atualizados. Algumas destas situações ocorrem às vezes pela junção de áreas de dois imóveis vizinhos no momento da oferta ao mercado. Estas situações foram analisadas individualmente e fez-se prevalecer às dimensões correspondentes ao valor do imóvel.

Os dados provenientes do ITBI foram analisados no sentido de identificar, principalmente, os valores subdeclarados, que normalmente ocorrem na cobrança desse imposto. Mas também de retirar alguns casos onde havia uma concentração de transações em imóveis com características semelhantes. Um exemplo disso são os novos loteamentos, que quando lançados à venda fazem com que haja um número considerável de transações.

Num primeiro momento a identificação dos casos em que poderiam se tratar de sub-declarações foi realizada pela comparação com os dados coletados de outras fontes. Para tanto, se definiram grupos de imóveis com características semelhantes pela utilização de técnicas de análise multivariada<sup>1</sup>, de modo a poder-se comparar os dados dentro de cada grupamento.

Num segundo momento, a identificação de casos atípicos dos dados de ITBI foi feita na própria modelagem<sup>2</sup>, onde se avaliavam os casos que apresentavam resíduos abaixo e acima de um determinado número de desvios padrões.

### 2.2. Modelagem

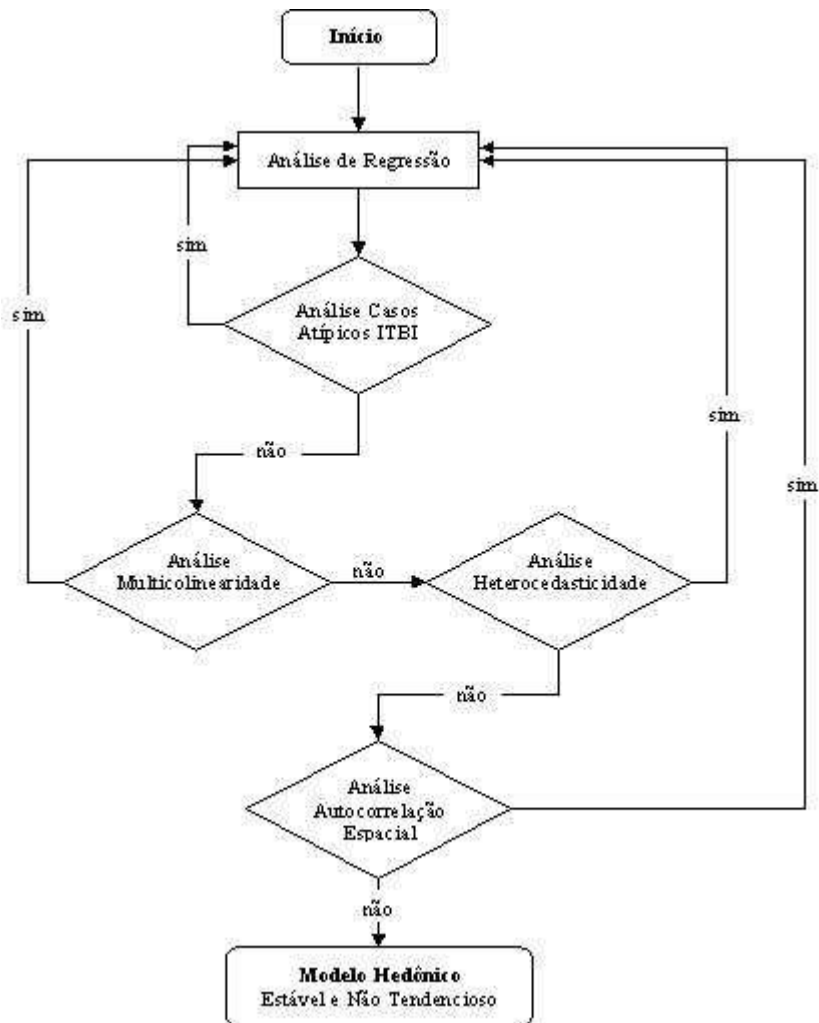
A modelagem dos dados foi realizada com o emprego da análise de regressão múltipla. Os

---

1 Análise Fatorial de Correspondências Múltiplas e Análise de Agrupamento.  
2 Análise de Regressão Múltipla.

atributos disponíveis no cadastro técnico, bem como aqueles gerados a partir do mesmo, foram as variáveis consideradas na análise. A modelagem teve dois objetivos. Num primeiro momento visou a identificação dos casos atípicos oriundos do ITBI, em razão de ser uma técnica que possibilita realizar suficientemente bem esta tarefa. Num segundo momento, o objetivo foi desenvolver um modelo que pudesse ser utilizado para a avaliação em massa dos terrenos existentes na área de estudo.

Inicialmente foram identificadas as variáveis que apresentavam melhores ajustes, para posteriormente se avaliar os casos atípicos. Os pressupostos básicos que a análise de regressão impõe foram observados com maior rigor quando da modelagem final, sendo os procedimentos utilizados para os processamentos seguidos conforme com o fluxo apresentado na FIGURA 1.



**Figura 1** : procedimentos para modelagem hedônica empregados no estudo.

### 3. Análise e Discussão

#### 3.1. Análise exploratória dos dados relativos ao mercado de terras

Os dados de mercado foram analisados inicialmente de forma exploratória, de maneira a buscar um primeiro entendimento das relações dos mesmos com as características que foram levantadas para cada um dos mesmos. Como visto anteriormente, os dados de mercado foram oriundos de duas fontes principais: 1) oferta (imobiliárias, corretores e proprietários) e; 2) transação (dados do ITBI). Havendo 299 e 1.239 eventos, respectivamente.

Para algumas características que serão utilizadas posteriormente no processo de modelagem fez-

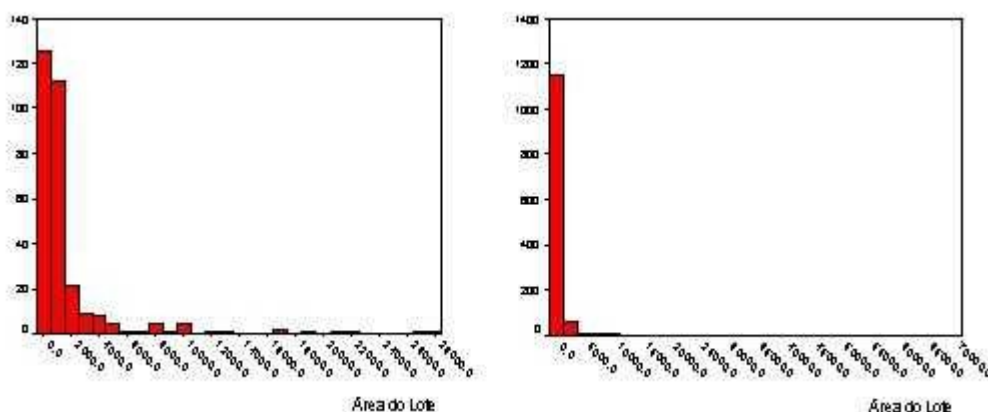
se a análise exploratória, onde se observou, principalmente, a variabilidade ou distribuição das mesmas, sendo os detalhes melhores vistos em SILVA (2006). A seguir, apresenta-se a análise da variabilidade realizada para a área e os valores global e unitário.

**TABELA 1** : limites mínimos e máximos da área e valores global e unitário.

Fonte	N	Área		Valor		VU	
		Mínima	Máxima	Mínimo	Máximo	Mínimo	Máximo
1- Oferta	299	166,50	28.235,90	10.000,00	1.600.000,00	2,02	1.501,50
2- Transação	1239	102,00	69.042,00	1.200,00	1.100.000,00	0,10	1.763,67

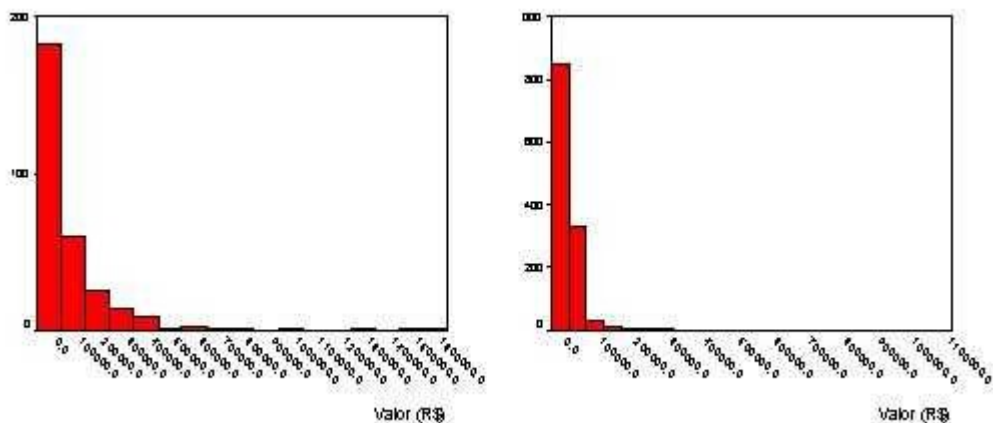
Observa-se pela TABELA 1 que o número de eventos oriundos do ITBI é muito superior ao que se conseguiu coletar nas fontes de oferta dos dados de mercado, representando 80,56% do total da amostra. Em relação aos limites mínimos e máximos demonstrados, nos chama mais atenção os valores máximos da área e os mínimos dos valores global e unitário, onde se percebe uma maior diferença. Quanto à área, há uma indicação de que os dados de oferta não conseguiram representar toda a amplitude da população, uma vez que o valor máximo das transações é consideravelmente superior e o da população muito mais (duas maiores áreas: 455.474,24 m<sup>2</sup> e 3.734.753,10m<sup>2</sup>). Como a pesquisa de mercado não conseguiu contemplar a totalidade das ofertas, assim como muitas transações ocorrem sem a exposição no mercado, é esperado que o campo amostral não tenha os mesmos limites dos dados da amostra de transações, que teve maior abrangência. Já, nos valores global e unitário, observa-se valores mínimos bem abaixo nas transações, o que de certo modo era esperado, em razão da subdeclaração dos valores dos imóveis que normalmente ocorre na cobrança desse tributo.

A seguir apresentam-se as freqüências relativas (histogramas) da área e dos valores.



**Figura 2** : distribuição da área do lote para os dados da amostra.

a) Freqüência relativa dados de oferta b) Freqüência relativa dados de transação



**Figura 3** : distribuição do valor para os dados da amostra.

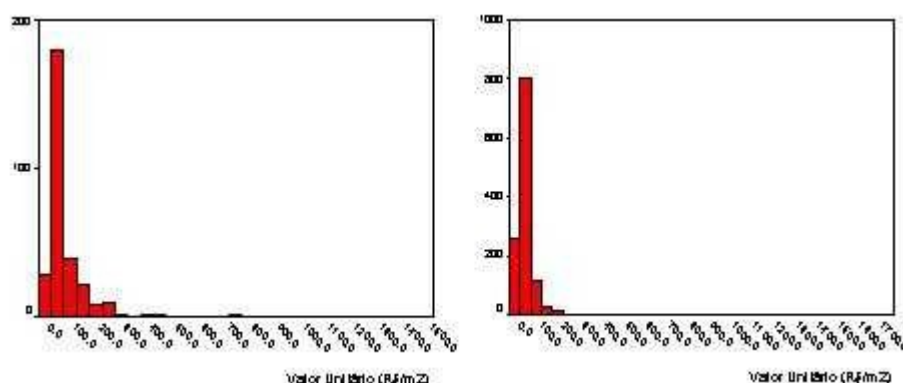
a) Freqüência relativa dados de oferta b) Freqüência relativa dados de transação

Pelas freqüências relativas apresentadas na FIGURA 2, observa-se que são poucos os eventos de mercado para as maiores áreas de lotes. Embora o limite máximo de área seja maior nos dados de transação, a freqüência de dados é praticamente nula no intervalo compreendido entre 10.000m<sup>2</sup> e 69.042m<sup>2</sup> de área. Comportamento não muito diferente para os dados de oferta. Destaca-se que as distribuições são de certo modo semelhantes, embora o número de dados não o seja, onde se tem uma maior freqüência para as áreas próximas dos lotes padrão, com um decrescimento rápido à medida que a área aumenta. Pode-se notar que as freqüências passam a ser baixas a partir de 5.000m<sup>2</sup> de área para ambas situações.

Este comportamento é esperado, haja vista que a proporção de lotes com áreas maiores tende a diminuir com o aumento da mesma, e, conseqüentemente, os eventos são igualmente menores.

As duas distribuições apresentadas na FIGURA 3 apresentam caudas bastante alongadas para direita, com os valores máximos muito distantes das faixas de valores com maior freqüência. No caso dos dados de transação o número de dados com valores acima de R\$200.000,00 é muito baixo. E para os dados de oferta a freqüência passa a ser muito baixa para os valores acima de R\$400.000,00. O cruzamento dos dois gráficos acima traz o indicativo que as ofertas existentes com valores superiores a R\$ 200.000,00 raramente constam na amostra das transações, diferentemente do que ocorre no mercado real, que opera nessa faixa de valor (acima de R\$ 200.000,00), deixando claro que esses valores dificilmente são capturados para efeito de ITBI (auto-declaração ou avaliação pelo município). Como os valores acima destes limites são comuns aos terrenos de maior área ou àqueles localizados na região central da cidade, pode-se concluir que nestas duas situações a amostra pode mostrar-se deficiente na modelagem estatística.

No caso dos valores unitários (FIGURA 4) a situação não é muito distinta, onde a freqüência de dados com valores unitários acima de R\$300,00 é baixa para ambas fontes. Para os dados de oferta ocorrem alguns dados acima deste valor até o limite máximo, mas que talvez não consigam contribuir para a modelagem estatística tendo o valor unitário como variável dependente.



**Figura 4** : distribuição do valor unitário para os dados da amostra.  
**a)** Freqüência relativa dados de oferta    **b)** Freqüência relativa dados de transação

No conjunto de dados amostrais foram encontrados 29 eventos de ofertas e 64 de transações duplos. Dos 29 casos de ofertas, um apresentava uma diferença relativa de 100%, podendo-se tratar de erro na coleta dos dados. Os demais apresentavam percentuais menores. Todavia, estas diferenças chamaram atenção. Nos 28 casos restantes a diferença média ficou em 16,53%. Estes eventos duplos foram de fontes ou épocas distintas.

É importante destacar que os valores ao longo do tempo tanto aumentaram quanto diminuíram. Em 15 casos a variação temporal foi menor que um ano; em 6 casos o valor aumentou com o tempo; e em 7 casos o valor diminuiu com o tempo. Logo, se concluiu que o tempo não influenciou significativamente a oscilação dos valores, devendo-se tratar de ajustes nas ofertas para melhor ajustar os valores que estavam sendo postos ao mercado.

Pôde-se perceber com isso, embora o número de casos seja pequeno, que existe uma margem de variação dos valores no mercado, não existindo um valor absoluto.

Em relação aos dados de ITBI, é de senso comum que os valores declarados normalmente apresentam alguma tendência, logo, fez-se necessário à confrontação dos mesmos com os dados de oferta para tentar-se identificar aqueles que apresentavam divergência significativa. Tomando em conta que a variabilidade é similar, entendeu-se que esta confrontação seria mais eficiente se fosse realizada por grupos com características homogêneas, sendo retirados da amostra 65 casos.

O procedimento empregado seguiu a metodologia apresentada por SILVA (1999), onde foram utilizadas técnicas de análise multivariada para definição dos grupos com características semelhantes. Os detalhes podem ser vistos em SILVA (2006).

Assim, considerando os 93 casos retirados por duplicidade e os 22 que estão acima de 10.000m<sup>2</sup>, eliminou-se com os procedimentos de análise exploratória 180 casos da amostra original, restando 1358 casos.

### 3.2. Modelagem do mercado de terras

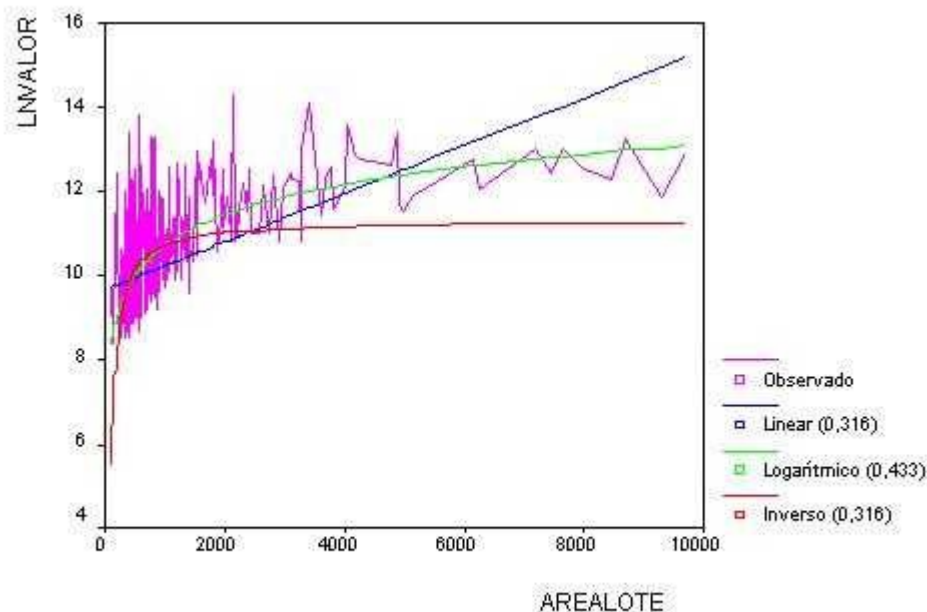
A modelagem do mercado de terras na área de estudo foi realizada com a utilização de análise de regressão múltipla. Num primeiro momento os processamentos tiveram o objetivo de se identificar os casos atípicos dos dados provenientes do ITBI, até se chegar no modelo que entendeu adequado para a inferência à população de imóveis representada pelo conjunto de dados amostrais. As variáveis que foram consideradas nos processamentos estão descritas na TABELA 2.

**TABELA 2** : variáveis utilizadas na análise de regressão.

	Variável	Tipo	Descrição	Forma de Entrada
1	VALOR	Contínua	Valor monetário do terreno	R\$
2	TESTADA	Contínua	Frente do lote para o logradouro principal	M
3	AREALOT	Contínua	Área do lote	M <sup>2</sup>
4	AREAED	Contínua	Área edificável	M <sup>2</sup>
5	GABARITO	Contínua	Largura do logradouro	M
6	POSQUA	Qualitativa	Posição do lote na quadra	1- meio e 2- esquina
7	ANO	Discreta	Ano do evento	1- 2003; 2- 2004 e 3- 2005
8	D_BIPTO	Contínua	Distância do imóvel ao centro da cidade	Km
9	DECLIVM	Contínua	Declividade média do lote	Graus decimais
10	COTAMIN	Contínua	Menor cota no lote	M
11	PERES	Contínua	Percentual de residências no trecho	%
12	AMRUA	Contínua	Área média das residências no trecho	Número de garagens
13	RENDA	Contínua	Renda média no trecho de logradouro	R\$
14	PAVIM	Qualitativa	Pavimentação	1- tem e 2- não tem
15	ELASTIC	Discreta	Elasticidade	1- oferta e 2- transação
16	TLOGRA	Discreta	Tipo do logradouro	1- residencial; 2- misto; 3- serviços

Cada uma das variáveis independentes (2 a 17) da TABELA 2 foi analisada inicialmente de maneira independente, de maneira a se verificar a relação das mesmas com a variável dependente. O gráfico apresentado na FIGURA 5 demonstra para a variável AREALOT o espalhamento dos pontos em um plano, com curvas de ajustes para se ter uma indicação da necessidade ou não de transformações para se obter um resultado mais satisfatório na modelagem final. A variável dependente demonstrou melhor ajuste quando transformada para logaritmo natural.

Pelo gráfico de espalhamento percebe-se, novamente, uma maior concentração dos lotes pesquisados nas áreas inferiores a 5.000m<sup>2</sup>. A área do lote quando transformada em logaritmo natural apresentou uma melhor correlação (0,433) com a variável dependente. Esta simples análise permite um entendimento das relações entre as variáveis, e foi realizada para todo o conjunto.



**Figura 5** : Relacionamento entre as variáveis AREALOTE e LNVALOR.

Após a análise descritiva iniciou-se a modelagem com o objetivo de identificar e retirar do conjunto de dados os eventos oriundos do ITBI que apresentavam valores abaixo dos praticados no mercado imobiliário. A cada resultado da análise de regressão, eram analisados os casos atípicos. Foram diversos processamentos, mas em resumo, o procedimento foi de avaliar os casos mais atípicos, abaixo de  $-5$  desvios padrão, e ir gradativamente reduzindo este número. Ou seja, após avaliar os casos com número de desvios abaixo de  $-5$ , por exemplo, fez-se novo processamento e novamente se avaliou os casos abaixo de  $-5$ , retirando os casos que se entendia tratarem-se de atípicos até que se pudesse concluir que os casos que restassem deveriam ser considerados nos processamentos subseqüentes. Daí, passou-se a avaliar os casos com números de desvio abaixo de  $-4$ , até chegar a  $-2$ . No decorrer dessas análises, os casos atípicos com desvios positivos também foram avaliados, sendo alguns casos também retirados. Por fim, foram retirados 477 casos de transação (ITBI) e 62 de ofertas.

Os resultados da modelagem que se entendeu mais apropriados são mostrados a seguir. Considerando-se as análises exploratórias e os resultados dos diversos processamentos realizados por análise de regressão, definiu-se o limite máximo de  $5.000\text{m}^2$  de área para os lotes que participariam da modelagem final, ficando a matriz composta de 999 linhas ou eventos. As variáveis contínuas que participaram da análise obtiveram na matriz final os seguintes resultados descritivos:

**Tabela 3** : estatísticas descritivas das variáveis contínuas.

Variáveis	N	Mínimo	Máximo	Média	Desvio	Curtose
LNVALOR	999	8.52	14.29	10.10	.82	2.91
TESTADA	999	1.00	90.00	17.90	9.01	16.40
GABARITO	999	4.00	30.00	12.12	4.30	5.63
D_BIPTO	999	.35	11.83	6.64	2.60	-.78
DECLINM	999	.21	30.18	9.48	6.18	-.22
AREALOT	999	166.50	4962.00	631.65	589.23	20.95
COTAMIN	999	2.94	193.18	52.26	28.76	2.24
PERES	999	5.51	100.00	89.51	18.15	3.65
AMRUA	999	68.00	462.40	169.02	48.62	4.71
RENDA	999	490.71	2918.16	1144.25	421.63	.82
AREAED	999	93.56	14346.39	999.85	1367.22	30.37

O valor pesquisado, como dito anteriormente, apresentou melhores resultados quando transformado em logaritmo natural. Sendo este tipo de constatação normalmente feita em estudos do mercado imobiliário, em razão, talvez, da dificuldade de se poder fazer um desenho amostral, pois não se pode definir previamente que imóveis seriam interessantes na composição da amostra, por se tratarem de eventos de mercado e que não estão sob o controle do pesquisador. Assim, dificilmente se encontra uma distribuição que seria ideal para a modelagem, tendo-se que recorrer a transformações que melhorem ou adequem o comportamento das variáveis para atender aos pressupostos das análises.

Na escala logarítmica o valor varia de 8,52 a 14,29, que correspondem na escala real a R\$5.000,00 e R\$1.600.000,00, respectivamente. Com média em 10,1047 ou R\$41.391,28.

Pelos coeficientes de curtose (*Kurtosis*) observa-se que as variáveis relacionadas às dimensões dos terrenos pesquisados são as que apresentam maior achatamento na curva de distribuição, enquanto que as demais tendem a ficar mais próximas de 0, que corresponde à distribuição normal. Quando se observa os quartis, nota-se que na maioria das variáveis a média está nos quartis inferiores, à exceção da variável PERES (percentual de residências no logradouro) que está no quartil superior. Assim, eventos com valores das variáveis que se situam nos quartis superiores tendem a se tornarem casos atípicos.

Quanto as variáveis qualitativas, destaca-se as variáveis TLOGRA (tipo de logradouro) e POSQUA (posição do lote na quadra) como as que apresentaram uma distribuição mais heterogênea, sobretudo a primeira. Situação que também não é muito apropriada para a modelagem. Todavia, se consideram importantes. Na variável POSQUA nada se podia fazer para deixar a distribuição mais homogênea, pois se trata da posição do lote na quadra. Já no tipo de logradouro, fez-se uma discretização, onde se poderia definir amplitudes que deixassem as freqüências mais homogêneas. No entanto, corria-se o risco de se juntar logradouros com importâncias muito distintas para o mercado imobiliário. A intenção foi, principalmente, destacar os logradouros da região central da cidade, no sentido de buscar um melhor ajuste para os eventos ali localizados.

As correlações parciais das variáveis independentes em relação à variável dependente apresentaram o comportamento esperado (sinal) e foram, de maneira geral, boas, à exceção da variável ANO que apresentou um índice baixo (0,095). Entre as variáveis independentes os índices se apresentaram na maioria dos casos em níveis baixos. Todavia, destacam-se três correlações acima de 0,7: 1- entre D\_BIPTO (distância ao centro) e RENDA; 2- entre AREALOT e AREAED (área edificável no lote) e; 3- entre PERES e TLOGRA. No primeiro caso, o índice foi de  $-0,872$ . De certo modo é um comportamento esperado, uma vez que as pessoas com poder aquisitivo mais elevado tendem a procurar a proximidade de facilidades urbanas, que normalmente estão nas regiões centrais da cidade. No segundo caso, a área edificável ou o potencial construtivo do lote depende em certo grau da dimensão do mesmo. Daí o índice mais elevado. E, no terceiro caso, a variável TLOGRA foi resultado de uma discretização da variável PERES, sendo esperado um comportamento mais estreito entre as duas.

Como a multicolinearidade acarreta em efeitos que desqualificam os modelos de regressão, fez-se uma combinação das variáveis que possuíam a mesma unidade, ou seja, somou-se os valores das variáveis AREALOT e AREAED, passando-se a denominar a nova variável de AREATOT. A mesma apresentou uma boa correlação com a variável dependente (0,647) e correlações parciais menos expressivas com as demais independentes.

Embora as correlações apresentadas anteriormente sejam consideradas elevadas para a modelagem por regressão múltipla, optou-se por manter as variáveis na análise e verificar se poderiam causar algum prejuízo aos resultados. Uma vez que se entendeu que são variáveis importantes no entendimento do comportamento do mercado imobiliário e que poderiam complementar a explicação da variabilidade dos valores.

A variável COTAMIN (menor cota no lote), quando processada junto com as demais variáveis na análise de regressão apresentou um comportamento difícil de ser explicado. Sua correlação com a variável dependente apresentou um sinal negativo, onde indicou que o valor do imóvel diminuiria à medida que sua cota ou altitude aumentasse. Talvez porque os terrenos de maior valor se encontram nas áreas mais baixas, como a região central da cidade. No modelo de regressão, seu sinal é positivo e apresenta-se como significativa. Com este sinal, a interpretação seria de que o valor do imóvel aumenta com sua altitude, deixando a dúvida quanto ao seu relacionamento com o valor do imóvel. Já, quando o processamento da análise de regressão foi realizado utilizando-se o método *Stepwise*, a mesma foi retirada do modelo, confirmando a existência de um comportamento difícil de ser explicado por este método de análise em um modelo geral, como este que se desenvolveu.

Deduz-se que a cota dos terrenos é um atributo importante na definição do valor dos imóveis, porém, ela apresenta diferentes influências em razão da localização dos mesmos. Em razão dos problemas de enchentes, terrenos situados em cotas mais altas tendem a ser valorizados, no entanto, a região central da cidade, que apresenta a maior valorização, situa-se numa região de cotas baixas.

Concluiu-se, em razão disso, que a altitude do imóvel interfere no valor segundo a localização do mesmo. Como apresentou índices de correlação acima de 0,3 com seis variáveis, destacando-se 0,639 com D\_BIPTO (distância ao centro) e -0,591 com a RENDA, retirou-se a mesma da análise por entender-se que as demais variáveis pudessem preencher razoavelmente a sua falta.

O modelo com 13 variáveis independentes teve um bom ajuste, considerando um modelo de avaliação em massa de imóveis e a sua área de abrangência, tendo um coeficiente de determinação da ordem de 82,9%, restando 17,1% de variabilidade não explicada. O coeficiente de correlação (0,910) demonstrou uma forte correlação entre a variável dependente e as independentes atuando conjuntamente. E o teste de significância do modelo (F de *Snedecor*) apontou para a rejeição da hipótese de não haver regressão a um nível de significância de 1%, com um índice calculado (366,49) bem acima do tabelado (2,13).

As variáveis, individualmente, foram testadas quanto às hipóteses de os valores calculados para t serem ou não diferentes de 0 (zero), a um nível de significância de 5% no teste t de *Student*. As variáveis que participam do modelo (TABELA 4) apresentaram os sinais esperados e compatíveis com os resultados correntes na literatura e trabalhos desenvolvidos nesta área, sendo a variável PAVIM (pavimentação) a única variável que ficou mais próximo do limite de significância pré-estabelecido, ficando o índice em 0,041, ligeiramente inferior a 0,05. Ainda que todas as variáveis tenham sido significativas, cabe ressaltar a afirmação de WONNACOTT e WONNACOTT (1981), onde explicam que o fato do coeficiente t (*Student*) de uma determinada variável  $X_j$  não ser discriminável a um nível de significância preestabelecido, não prova a falta de relacionamento entre esta e a variável dependente. Quando se têm fortes razões teóricas para acreditar que Y é positivamente ou negativamente explicado por  $X_j$ , que pode ser confirmado pelo sinal do coeficiente; embora se tenha uma fraca confirmação estatística, é prudente rejeitar a hipótese nula ( $H_0: b_j = 0$ ).

**Tabela 4** : modelo de regressão múltipla.

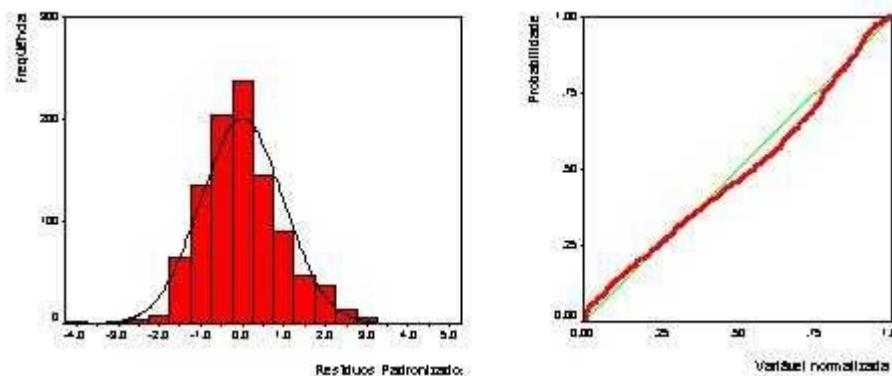
Variáveis	Coeficientes não Padronizados		t	Sig.	95% de Intervalo de Confiança para B		Estatísticas de Colinearidade	
	B	Erro Padrão			Limite Inferior	Limite Superior	Tolerância	VIF
(Constant)	9.538	.211	45.265	.000	9.124	9.951		
TESTADA	7.846E-03	.001	5.270	.000	.005	.010	.679	1.473
GABARITO	1.843E-02	.003	4.805	.000	.010	.023	.537	1.861
AND	7.477E-02	.015	4.952	.000	.045	.104	.991	1.009
POSQUA	8.844E-02	.033	2.709	.007	.024	.153	.943	1.061
D_BIPTO	-5.096E-02	.009	-5.611	.000	-.069	-.033	.207	4.820
DECLIVM	-1.036E-02	.002	-5.636	.000	-.014	-.007	.898	1.114
PERES	-4.007E-03	.001	-3.764	.000	-.006	-.002	.311	3.215
AMRUA	1.427E-03	.000	4.964	.000	.001	.002	.594	1.683
RENDA	5.945E-04	.000	10.958	.000	.000	.001	.222	4.508
PAVIM	-5.807E-02	.028	-2.047	.041	-.114	-.002	.649	1.540
ELASTIC	-.298	.027	-11.210	.000	-.350	-.246	.906	1.104
TLOGRA	.221	.058	3.785	.000	.106	.335	.408	2.451
AREATOT	1.391E-04	.000	18.397	.000	.000	.000	.566	1.767

As variáveis independentes entraram todas na forma direta, sem nenhum tipo de transformação. Os valores de VIF (*variance inflation factor*) estão abaixo do limite máximo, que segundo DES ROSIERS *et al.* (2003) é de 10, onde o maior valor foi de 4,820 para a variável D\_BIPTO (distância ao centro).

Os resíduos foram investigados ainda quanto à normalidade, podendo-se afirmar que não houve violação desse pressuposto, pois pelas propriedades da curva normal: 68% dos resíduos devem estar no intervalo (-1; 1), 90% no intervalo (-1,64; 1,64) e 95% no intervalo (-1,96; 1,96); e a distribuição apresentada pelos resíduos nesta análise foram os seguintes: 71,9% no intervalo (-1; 1), 91,3% no

intervalo (-1,64; 1,64) e 94,9% no intervalo (-1,96; 1,96). Sendo percentuais próximos dos pré-estabelecidos para curva normal.

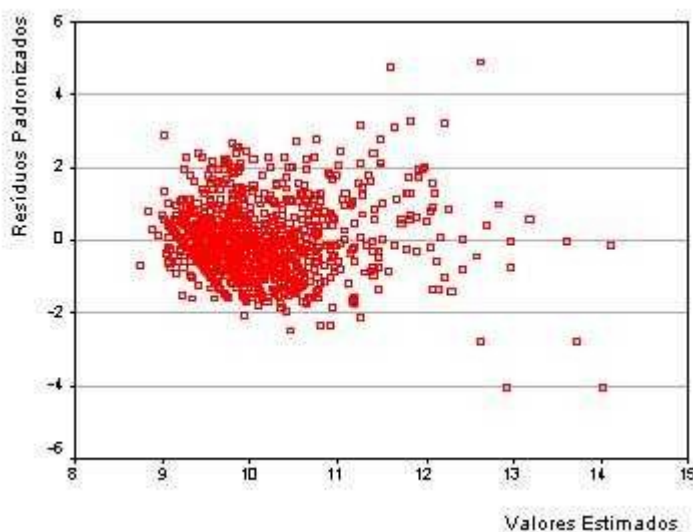
Os gráficos a seguir apresentam a freqüência relativa e o gráfico de probabilidade normal dos resíduos.



**Figura 6** : freqüência relativa e gráfico de probabilidade normal dos resíduos.  
a) Freqüência relativa b) Gráfico de probabilidade normal

Os gráficos (FIGURA 6) confirmam o que foi dito anteriormente, onde pode-se observar visualmente que não existem fugas da distribuição normal para os resíduos da regressão. Quando se observa a relação dos resíduos com os valores estimados para os dados amostrais (FIGURA 7), percebe-se que os resíduos apresentam uma distribuição que se pode considerar aleatória. No entanto, com um pequeno número de casos com resíduos maiores associados a valores estimados mais elevados. O que não se entendeu como pontos que pudessem perturbar de forma significativa o ajuste do modelo.

Pode-se observar, ainda, pelo gráfico da FIGURA 44 os pontos que possuem resíduos mais elevados. Após a padronização dos resíduos, de maneira a se identificar os casos atípicos, destacaram-se com resíduos acima de 2 desvios padrão 39 casos, e desses, 6 casos estão acima de 3 desvios. Alguns autores de trabalhos relacionados a avaliação em massa de imóveis acreditam que a análise de pontos atípicos não deve se restringir ao exame do limite de dois desvios-padrão, uma vez que tratam de explicar um comportamento muito mais heterogêneo do que as avaliações individuais. Todavia, é importante enfatizar que o fato de se permitir um maior relaxamento, não signifique que deixe de ser necessário investigar se os resíduos mais elevados podem ser fonte de perturbações no modelo.

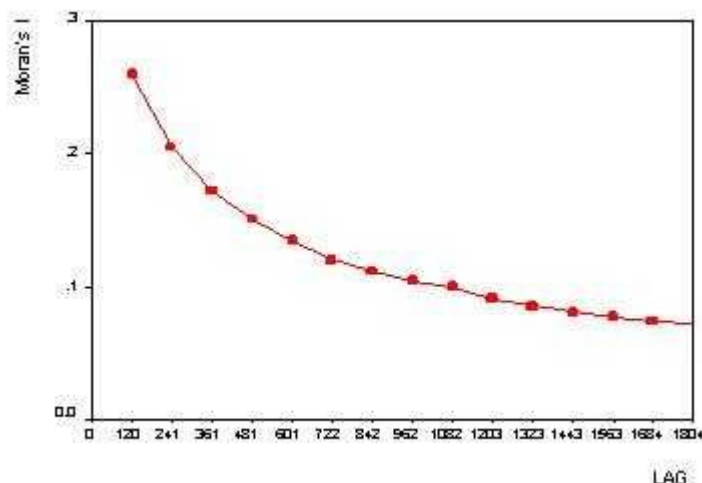


**Figura 7** : Resíduos padronizados x valores estimados da regressão.

Em relação a dependência espacial dos resíduos, calculou-se o índice de Moran para para diferentes distâncias entre pares de pontos (LAG), iniciando em 120m, onde para este LAG o índice foi de 0,26. Este valor demonstra, embora seja um índice baixo, que ainda existe uma pequena dependência espacial dos resíduos, que muito provavelmente permanece em razão de ausência de variáveis que possam complementar a variabilidade não explicada da localização. Evidentemente que não se pode atribuir este peso unicamente para localização. Outros fatores podem estar contribuindo, como a falha na medida de determinadas variáveis ou perda de informação em razão de análises prévias, bem como valores de determinados imóveis com alguma tendência ou mesmo erros, por exemplo.

Os índices de autocorrelação espacial da variável dependente (LNVALOR) e do valor estimado pelo modelo de regressão foram de 0,50 e 0,53, respectivamente, para o mesmo LAG inicial descrito anteriormente (120m). Estes valores indicam uma dependência espacial significativa dos valores de mercado, e que, pelo grau de dependência espacial dos resíduos, os atributos de localização considerados no modelo deram conta de boa parte da variabilidade da componente espacial do valor.

A FIGURA 8 a seguir apresenta o Correlograma de Moran para os resíduos da regressão, onde se percebe um decréscimo a partir do LAG inicial, com índices inferiores a 0,1 a partir de 1.082m de distância entre os pares de pontos resíduos.



**Figura 8** : correlograma de Moran para os resíduos da regressão.

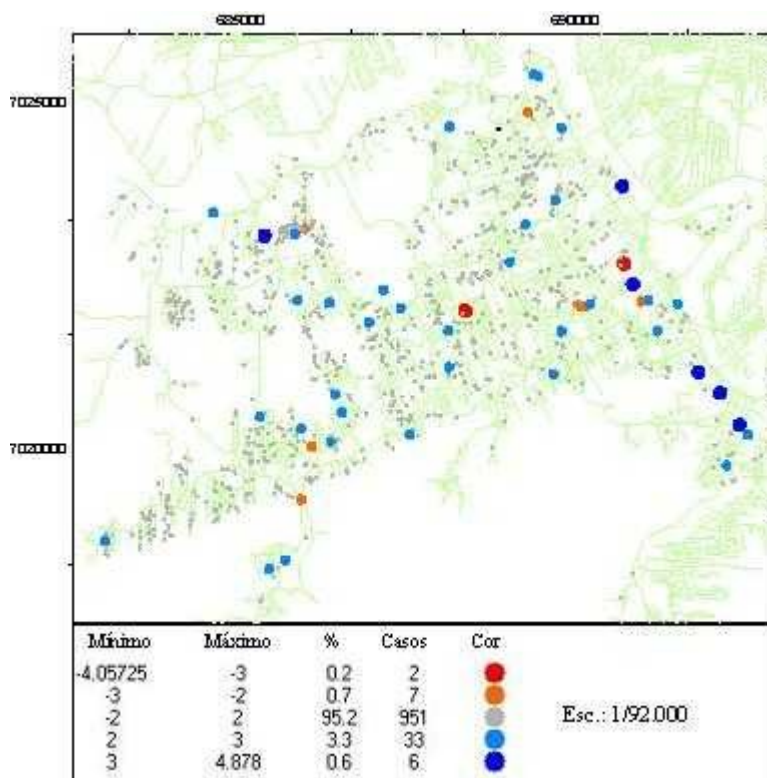
Os resíduos padronizados foram espacializados por pontos em cores que representam intervalos, de modo a poder se perceber o comportamento dos mesmos (FIGURA 9). Os maiores resíduos foram representados com cores mais intensas e em círculos maiores. Os que ficaram entre  $-2$  e  $2$  desvios padrão estão representados por círculos menores de em cinza claro. A classificação da distribuição dos resíduos (e) é a seguinte: 1)  $-3 > e$ ; 2)  $-2 > e \geq -3$ ; 3)  $2 \geq e \geq -2$ ; 4)  $2 < e \leq 3$ ; 5)  $3 > e$ .

**TABELA 5** : dados cruzados entre elasticidade e classes de resíduos padronizados.

Elasticidade	Distribuição dos Resíduos					Total
	1	2	3	4	5	
1- Oferta	-	5	217	10	5	237
2- Transação	2	2	734	23	1	762
Total	2	7	951	33	6	999

Na TABELA 5 tem-se o cruzamento das freqüências dos resíduos padronizados segundo os intervalos descritos anteriormente com a elasticidade dos dados.

Percebe-se pelos números apresentados na TABELA 5 e pela visualização da FIGURA 9 que nos intervalos de resíduos maiores (1, 2, 4 e 5) há um predomínio dos que são positivos. Isso acontece, talvez em parte, pelo fato de que se preocupou, na análise exploratória dos dados, em se retirar do conjunto de dados as subdeclarações do ITBI. Retirando-se poucos casos com resíduos positivos maiores. Por outra parte, quando se observam os resíduos da classe 5 ou maiores que 3 desvios padrão, verifica-se que dos 6 casos, 5 são de ofertas. Isto se deve provavelmente ao peso ou quantidade de dados de transação na amostra, que tendem a ter valores menores do que os de oferta, e pela menor representatividade dos eventos de maior valores que normalmente ocorrem na região central da cidade, mas que tendem a apresentar um maior grau de elasticidade. Pela FIGURA 9 pode-se ver que três casos encontram-se no centro da cidade, dois próximos ao centro e 1 mais afastado.



**Figura 9** : Distribuição espacial das classes de resíduos padronizados.

Pelos resultados apresentados na análise do modelo de regressão múltipla, pôde-se concluir que o mesmo apresenta condições de ser utilizado para estimação de valores para a massa de terrenos compreendidos nos intervalos de variação das variáveis consideradas. A sua aplicação para terrenos de maior valor localizados na região central da cidade fica um pouco limitada, uma vez que na estimação desses casos na amostra demonstrou que a estimação tende a ser consideravelmente menor que a observação. Devendo seu emprego ser dado com cautela para esta situação.

Outras situações podem fugir ao controle do modelo, como por exemplo, terrenos encravados (sem frente imediata para o logradouro de acesso) abaixo do nível do logradouro. Como o número de casos amostrado foi muito pequeno, não foi possível definir variáveis ou pesos que pudessem contribuir para a explicação deste comportamento, sendo os mesmos retirados da amostra.

Como foi um modelo ajustado para avaliação em massa de terrenos baldios, que na área de estudo representam 21,51% dos terrenos existentes, e ainda considerando as limitações apresentadas nos parágrafos anteriores, sua utilização pode ser consideravelmente limitada se fosse eleito como modelo para definição dos valores venais ou base de cálculo para os tributos imobiliários. Todavia, se seu emprego fosse para definir valores unitários por trecho de logradouro ou face de quadra no sentido de obter uma planta de valores genéricos, seu potencial de uso seria aumentado enormemente.

Das variáveis estudadas e aproveitadas no modelo de regressão, as de maior importância são: RENDA (renda média), ELASTIC (elasticidade do evento) e AREATOT (área do lote mais a área edificável), conforme indicado através da estatística de *Student*.

Fica evidenciada no modelo de regressão, especialmente por conta das variáveis mais importantes, a relação do valor da terra com a capacidade de aproveitamento e de geração de riqueza sobre a mesma.

Para o caso de tributação imobiliária, que se baseia no valor venal do imóvel, o presente trabalho serve para demonstrar a relação entre a riqueza (valor da terra) e a capacidade contributiva (importância da variável renda no modelo) do proprietário.

#### 4. Considerações Finais

A modelagem do mercado de terras desenvolvida nesse trabalho teve como objetivo evidenciar que é possível qualificar os procedimentos de avaliação nas administrações municipais, com base em um cadastro técnico estruturado de maneira contemplar dados úteis a este fim, bem como para facilitar o uso desses dados. Neste contexto, entende-se que os trabalhos relacionados à manutenção de um banco de dados do mercado imobiliário devam ser uma componente que permita tratar, de maneira qualificada, um dos aspectos básicos do cadastro, que é o econômico.

O conjunto de dados disponíveis sobre as propriedades possibilitaram evidenciar atributos importante na explicação do comportamento do mercado de terras na área de estudo. Todavia, em razão da heterogeneidade desse tipo de imóvel tratado pela modelagem, percebeu-se que o conjunto de eventos de mercado pesquisados limitou o ajuste do modelo a parte dos imóveis.

Apesar da heterogeneidade e a limitação do uso de determinadas variáveis por apresentarem colinearidade ou um comportamento difícil de ser captado por um modelo de regressão, pode-se dizer que os resultados foram bons, diante dos resultados de pesquisas semelhantes. Melhores resultados poderiam ter sido obtidos se se dispusesse de um conjunto maior de eventos de mercado, dando possibilidades de trabalhar com desenho amostral, e assim atender melhor aos pressupostos básicos da análise de regressão. Outras técnicas de análise de dados podem muito bem complementar ou mesmo substituir a análise de regressão, porém, todas necessitam, da mesma forma, de um maior número de casos possíveis para terem efetividade.

A aplicação do modelo para definição de valores unitários de referência para uma planta de valores é viável, bem como a adequação de fatores para a composição do modelo a ser utilizado para a massa dos imóveis.

#### 4. Referências Bibliográficas

- DES ROSIERS, F.; THÉRIAULT, M.; MÉNÉTRIER, L. D..** *Spatial versus non-spatial determinants of shopping center rents : modeling location and neighborhood-related factors*. Document de Travail, Faculté des Sciences de L'Administration, Université Laval, Québec, 2003-18, 32 p.. Disponível : <http://www.fsa.ulaval.ca/rd>.
- SILVA, E..** *Proposta de avaliação coletiva de imóveis : aplicação aos imóveis do tipo apartamento na cidade de Blumenau – Santa Catarina*. Florianópolis, 1999. Dissertação de mestrado em engenharia civil, Universidade Federal de Santa Catarina – UFSC, 95 p..
- SILVA, E.** *Cadastro técnico multifinalitário: base fundamental para avaliação em massa de imóveis*. Florianópolis, 2006. Tese de pós-graduação em engenharia de produção, Universidade Federal de Santa Catarina – UFSC.