

# Modelagem empírica de rentabilidade no mercado de locação de imóveis na cidade de São Paulo/SP utilizando modelos hedônicos e de regressão

Fábio Gerab <sup>[1]</sup>, Bruno Mariano de Carvalho <sup>[2]</sup>

[1] [prifgerab@fei.edu.br](mailto:prifgerab@fei.edu.br). Centro Universitário FEI / Departamento de Matemática. [2] [bruno.mariano@hotmail.com](mailto:bruno.mariano@hotmail.com). Centro Universitário FEI / Departamento de Engenharia Mecânica.

## RESUMO

É um grande desafio determinar o valor intrínseco associado a todo o pacote de atributos existentes em um imóvel e prever o seu valor da maneira eficaz, pois, no mercado imobiliário, existem diversos imóveis com características distintas. Este trabalho tem como objetivo desenvolver um modelo matemático capaz de estimar o retorno financeiro de investimentos em apartamentos residenciais urbanos na cidade de São Paulo/SP, destinados à locação, identificando as características que agregam maior rentabilidade para estes apartamentos residenciais. Para isso, modelos de precificação hedônica, capazes de prever valores de venda e de aluguel, foram desenvolvidos. Através destes modelos e da razão entre os valores preditos de locação e de venda dos imóveis da amostra, foi possível construir um novo modelo de regressão para a rentabilidade na locação e determinar quais são as características presentes em apartamentos residenciais que apresentam maior rentabilidade do investimento.

**Palavras-chave:** Mercado imobiliário. Modelos hedônicos. Regressão linear múltipla. Rentabilidade.

## ABSTRACT

*It is a big challenge to determine the intrinsic value associated with all attributes existing in a property. It is also difficult to predict this value accurately, because in real estate market there are several properties with distinct characteristics. This study aims at developing a mathematical model capable of estimating the financial return on investments in urban residential apartments destined to lease in São Paulo city, SP, and identifying characteristics that promote a greater profitability for these residential apartments. In order to do so, hedonic multiple linear regression models capable to predict the values of sale and rent have been developed. With these models, and the ratio between the predict values of rent and sale, it was possible to build a new regression model in order to determine the profitability in the rent of residential apartments. Using the analysis of this new model it was possible to determine the characteristics on apartments that present the best profitability on real estate market.*

**Keywords:** Hedonic models; Multiple linear regression; Profitability, Real estate market.

## 1 Introdução

O mercado imobiliário é um mercado de suma importância para o crescimento de uma nação, pois está relacionado não só à comercialização de imóveis mas também com a construção e sua ocupação.

As comercializações no mercado imobiliário são compostas por transações de bens imobiliários, por meio das quais imóveis são adquiridos, vendidos ou alugados. Para o funcionamento dessas transações, no mercado imobiliário devem existir três componentes fundamentais: os vendedores, os compradores e os imóveis (GAZOLA, 2002).

Para que haja transações de bens imobiliários é necessário que estes possuam um valor de comercialização atribuído, este por consequência surge a partir de uma avaliação. A avaliação de imóveis consiste na determinação do valor de mercado de um imóvel, ou seja, o preço mais provável que o bem pode atingir em uma transação, dada as suas características e condições de mercado (MOREIRA; SILVA; FERNANDES, 2010).

Devido a inúmeras divergências de características dos imóveis, os preços no mercado imobiliário tendem a um comportamento diferenciado. Além do dinamismo intrínseco dos processos de estruturação interna nas áreas urbanas, os governos e a economia exercem grande influência na formação dos preços (BRAULIO, 2005). Características físicas, sociais e naturais, entretanto, também podem influenciar na definição dos valores dos imóveis. Outra influência tem a ver com a valorização ou desvalorização sofrida pelo mercado imobiliário, que por sua vez está relacionada com a lei da oferta e da procura (SANTELLO, 2004).

Partindo da premissa que todos os aspectos apresentados são possíveis influenciadores no preço dos imóveis, faz-se necessário uma avaliação do bem, verificando suas características intrínsecas e extrínsecas (ALVES, 2005). Os métodos estatísticos são os procedimentos de modelagem matemática mais utilizados para a avaliação imobiliária (BAPTISTELLA, 2005), pois estes possuem poder de análise em grande escala. Devido à complexidade destes métodos utilizados na avaliação, seu uso fica restrito a profissionais altamente capacitados, não se difundindo entre os avaliadores que atuam no mercado (BRONDINO; SILVA, 2005).

A preocupação, por parte do governo e outros órgãos, com o aumento populacional em grandes cidades, como São Paulo/SP, torna o mercado

imobiliário tema de diversas discussões entre artigos acadêmicos visando sua melhor compreensão, pois a procura por um bem imobiliário passa a aumentar juntamente com este aumento populacional, devido à necessidade básica de moradia de qualquer indivíduo.

Investimentos no mercado imobiliário possuem um risco menor quando comparados com investimentos de alto risco, como, por exemplo, a bolsa de valores. Devido à característica do imóvel de ser um bem concreto, o risco agregado a ele é menor. Este risco atribuído ao investimento imobiliário é referente à incerteza de qual e quando será o retorno financeiro do investimento; também diz respeito à desvalorização que pode ocorrer no mercado imobiliário. Por essas razões, o investimento no mercado imobiliário pode ser definido como de risco moderado.

Existindo procura e oferta, risco moderado no investimento e grande número de transações de bens presentes no mercado imobiliário, os investidores procuram tal investimento em busca de adquirir um imóvel para alugar ou vender posteriormente.

Devido à grande quantidade de atributos existentes em cada imóvel, é possível se afirmar que, dentro do contexto econômico, o mercado imobiliário apresenta uma série de características complexas que dificultam sua análise com a devida precisão.

Por fim, é de interesse de quem busca investir no mercado imobiliário, neste caso em apartamentos residenciais urbanos, saber quais características de imóveis geram maior rentabilidade.

## 2 Fundamentação teórica

### 2.1 Modelos hedônicos

Para se entender melhor o que são modelos hedônicos, primeiro é preciso um entendimento do que é hedonismo. Trata-se do prazer no uso ou consumo de um bem. Negri Neto (1976) descreve hedonismo como uma doutrina que considera o prazer individual e imediato como o único bem possível. De maneira geral são as doutrinas que concordam na determinação do prazer como bem supremo, dedicação ao prazer do sentido e busca pelo prazer.

Ao adquirir um bem, a pessoa busca não apenas o objeto em si mas também a utilidade ou o prazer que este pode proporcionar a ela. O preço do bem, então, é formado pelas características pertencentes a ele, e a utilidade que estas características presentes no

bem trazem à pessoa que o adquire. Esta combinação forma o seu preço hedônico.

De acordo com Rosen (1974), preços hedônicos são definidos como sendo os preços implícitos dos atributos e são revelados a partir de preços observados de produtos diferenciados e das respectivas características a eles associadas. Considera-se, desta forma, que existe um conjunto de características habitacionais que podem explicar o valor do patrimônio.

Para se determinar o valor de um bem, utiliza-se o preço existente do bem no mercado, mas, devido a situações nas quais há inexistência de um bem no mercado para fazer a comparação, pode-se recorrer a valores indiretos que se aproximariam do preço do bem, desmembrando o valor do bem em valor de uso e valor de não uso, os quais, por sua vez, seriam, respectivamente, o valor referente ao uso efetivo ou potencial que o bem pode prover (valor de uso) e o valor intrínseco que existe no bem (valor de não uso) (NEGRI NETO, 1976).

Os preços hedônicos estão presentes em diversas áreas do comércio onde o preço do bem é a utilidade do conjunto de seus atributos. Desta forma, onde existam características distintas em distintos bens, como no mercado imobiliário, tende a existir a presença de preços hedônicos. Braulio (2005) diz que os preços dos imóveis podem ser compreendidos como a soma dos produtos das quantidades de cada um desses atributos pelos seus preços implícitos, ou hedônicos.

A teoria clássica do consumidor muito tem a ver com o preço hedônico em bens de consumo, já que esta busca descrever como os consumidores tomam decisões de compra entre outros aspectos abrangentes. Esta teoria apresenta deficiências na definição da relação entre os bens (ANGELO; FÁVERO, 2003).

Os preços hedônicos existentes em bens de consumo podem ser quantificados com a utilização de modelos matemáticos. O princípio básico de modelos de preços hedônicos decorre de uma nova abordagem da teoria clássica do consumidor denominada teoria do consumo de Lancaster (ANGELO; FÁVERO, 2003). A teoria do consumo de Lancaster cita que a utilidade de um bem deriva das suas propriedades e ou características (MARQUES; CASTRO; BHATTACHARJEE, 2009).

Os Modelos de preços hedônicos tipicamente utilizam de análise de regressão linear múltipla. Segundo Fávero (2003), estes modelos têm sido utilizados para se avaliarem as características de

imóveis no mercado imobiliário, pois se referem a valores implícitos das características dos imóveis.

No mercado imobiliário, os modelos de preços hedônicos são utilizados a partir da regressão, ou seja, o preço de venda dos imóveis é regredido por suas características, gerando-se, assim, o valor de mercado de suas características.

Em cada imóvel existe uma grande quantidade de características que o definem – não apenas as características internas mas também as externas. As características externas, ou externalidades, presentes no imóvel influenciam diretamente o seu preço de mercado. Alguns exemplos de externalidades são: número médio de pessoas por domicílio; acesso à água e esgoto; entre outros (CARMO, 2014).

As externalidades estão diretamente relacionadas com o preço hedônico atribuído ao imóvel. Abramo e Faria (1998) afirmam que as externalidades exercem significativo efeito sobre o preço do imóvel, devendo ser consideradas como suas características implícitas. Desta forma, as características externas do imóvel têm significância na formação do preço atribuído ao imóvel, sendo que o consumidor busca não apenas o conjunto de atributos presentes no imóvel mas também o conjunto de atributos externos que podem ter relevância para quem o adquire.

Segundo Fávero (2003), a observação das características externas é recomendada, pois as variações nos preços de venda dependem da localidade do imóvel, logo, dependendo da localidade do imóvel, o valor de suas características tende a sofrer variação.

## 2.2 Regressão linear múltipla

A regressão linear múltipla é usada para estudar a relação entre uma variável dependente e várias variáveis independentes explicativas, conhecendo-se e controlando-se as variáveis que contribuem para a ocorrência do fenômeno.

A Equação 1, que representa a regressão linear múltipla (GUERRA; DONAIRE, 1991), é dada por:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + U \quad (1)$$

Em que:

$\alpha$  = constante da regressão;

$\beta_i$  = coeficientes angulares de regressão;

U = variável aleatória do fenômeno (erro aleatório).

Existem alguns conceitos que devem ser analisados em relação aos modelos de regressão utilizados neste estudo, que são: Outliers, a autocorrelação e multicolinearidade e a homocedasticidade, descritos nesta ordem, abaixo:

Outliers, ou pontos influenciadores, são elementos que se comportam de maneira atípica aos demais elementos existentes na pesquisa, trazendo perturbações ao modelo de regressão.

Não deve existir autocorrelação de erros e multicolinearidade entre as variáveis do modelo, pois variáveis que possuem autocorrelação de erros e multicolinearidade tendem a perturbar o modelo de regressão.

A condição da homocedasticidade, ou variância constante, é de fundamental existência nos modelos de regressão linear múltipla.

### 2.2.1 Poder de explicação dos modelos de regressão

Para verificar quão ajustado está o modelo, calcula-se o coeficiente  $R^2$ , chamado de coeficiente de determinação múltipla, que indica o poder de explicação do modelo em função das variáveis independentes. A expressão é dada pela Equação 2:

$$R^2_{ajustado} = 1 - (1 - R^2) \times \frac{n-1}{n-k-1} \quad (2)$$

Em que:  $Y_i$  = valores observados;  $\langle y \rangle$  = média dos valores observados;  $y_i$  = valores estimados pela equação de regressão.

A desvantagem deste coeficiente é que, para uma mesma amostra, ele cresce na medida em que aumenta o número de variáveis independentes incluídas no modelo, não levando em consideração o número de graus de liberdade perdidos a cada parâmetro estimado (DANTAS, 1998). Alguns autores recomendam o uso do coeficiente de determinação ajustado, que leva em conta o número de variáveis explicatórias em relação ao número de observações. A expressão para o coeficiente de determinação ajustado, Equação 3, é dada por:

$$R^2 = \frac{\sum_{i=1}^n (y_i - \langle y \rangle)^2}{\sum_{i=1}^n (Y_i - \langle y \rangle)^2} \quad (3)$$

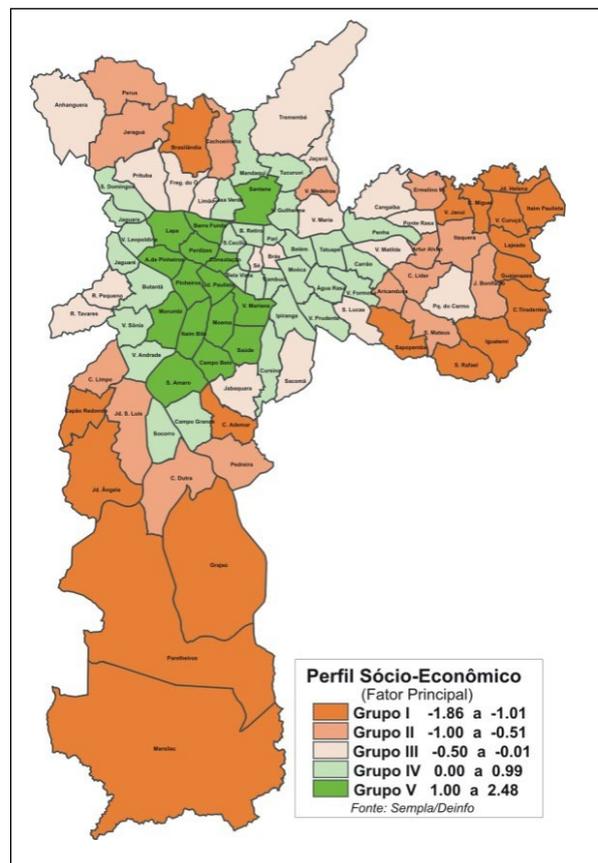
em que: R = coeficiente de determinação; n = número de elementos da amostra; k = número de variáveis independentes do modelo.

## 3 Metodologia

A metodologia deste trabalho utiliza os índices de distribuição socioeconômica dos distritos de São Paulo/SP (Atlas Ambiental do Município de São Paulo) e as informações disponibilizadas em anúncios de imóveis, por intermédio do site ZAP Imóveis, para a construção de uma base de dados.

A amostragem foi feita utilizando-se o índice socioeconômico dos distritos localizados em São Paulo, com ênfase nas diferentes faixas de distribuição de renda, como demonstra a Figura 1 abaixo (SÃO PAULO, 2016). Desta forma a amostra foi dividida em 3 grupos – V, III e I+II, respectivamente, maior índice, índice mediano e o menor índice. A coleta dos dados foi feita tanto para locação quanto para venda de imóveis.

Figura 1 – Mapa perfil socioeconômico do município de São Paulo



Fonte: Sempla/Deinfo

Outra estratégia adotada para a amostragem foi a de construir a amostra tendo o número de dormitórios existentes no imóvel como a característica de relevância; assim dividiu-se a amostra entre imóveis de 1, 2, 3 e 4 ou mais dormitórios. O objetivo de tal estratégia foi estratificar os imóveis, pois, como avaliado a priori, imóveis com quantidades diferentes de dormitórios tendem a possuir características espaciais distintas.

### 3.1 Características amostrais

O tamanho da amostra foi definido de maneira que, para cada grupo, foi escolhido um bairro do qual foram coletadas informações de 200 imóveis, 100 imóveis à venda e 100 sendo locados, totalizando 600 dados de imóveis coletados. Devido à menor quantidade de dados disponíveis para os grupos I e II, associados ao perfil socioeconômico mais modesto, foram utilizados diversos bairros de faixas de renda semelhantes para a amostragem dos dados destes grupos.

Assim, os bairros selecionados para a amostragem foram:

**Grupo V** – Moema

**Grupo III** – Bela Vista

**Grupo I e II** – Itaquera, São Mateus, Guaianases, Cidade Tiradentes, Aricanduva, Cangaíba, Cidade Líder, Artur Alvim, Lajeado, Vila Curuçá, Sapopemba, Itaim Paulista, São Miguel Paulista, Brasilândia e Vila Jacuí.

Com o objetivo de obter maior imparcialidade na construção da amostra, visando diminuir interferências externas, a amostragem foi realizada de forma pseudoaleatória, utilizando um gerador de números aleatórios para a seleção dos imóveis a serem incluídos na amostra.

Para definir quais imóveis seriam coletados, foi inicialmente verificada a quantidade de anúncios de imóveis existentes no site ZAP Imóveis na época da amostragem, para a região escolhida. Em seguida, verificou-se a quantidade de anúncios de imóveis existentes para cada número de dormitórios. Estes dados foram computados e, utilizando-se uma amostragem proporcional, a quantidade de imóveis coletados foi definida de acordo com a proporção dos imóveis anunciados para cada número de dormitórios.

Após ter a proporção definida, foi verificada a quantidade de páginas de anúncios existentes para cada número de dormitórios e, se o número de páginas fosse superior ao número de imóveis necessários para a amostra, se utilizava o gerador de números aleatórios

para definir as páginas das quais os imóveis seriam coletados. Na sequência era determinado, também se utilizando um gerador de números aleatórios, o número de ordem na página do anúncio a ser selecionado para a amostra.

Caso o número de páginas fosse menor que o número de imóveis a serem coletados, o processo acima era repetido até que se completasse a amostra, tendo-se o cuidado de se garantir a impossibilidade de que um mesmo imóvel fosse inserido na amostra em duplicidade. Desta forma o processo de amostragem foi probabilístico, tomando como população o conjunto dos imóveis (apartamentos residenciais nos bairros selecionados) anunciados.

Para cada imóvel selecionado, encontrado no site do ZAP Imóveis, foram coletadas todas as informações disponibilizadas utilizando escalas ad hoc para este processo.

As variáveis coletadas para a análise foram: N° Dorm. (número de dormitórios), Preço (preço do imóvel para venda ou para aluguel em R\$), Área Útil (em m<sup>2</sup>), Vagas (número de vagas), Condomínio (valor do condomínio), Portaria 24h, Mobiliado (apartamento mobiliado), Área de Serviço, Arm. Cozinha (armário da cozinha), Arm. Embutido (armário embutido), Ar Condicionado, Brinquedoteca, Churrasqueira, Copa, Quadra, Depósito, Elevador, Esp. Gourmet (espaço gourmet), Academia, Lavabo, Sauna, Piscina, Playground, Quarto Empregada (quarto de empregada), Salão de Festa, Sala de Jogos, Sacada, Interfone, Hidromassagem, ano de construção, IPTU (valor do IPTU) e Lavanderia Externa.

A escala ad hoc foi definida a partir da análise da amostra, de forma que:

- caso a variável não existisse no imóvel, se computava 0 (zero), certeza da ausência da variável;
- caso não houvesse certeza da existência da variável, mas as características do imóvel indicassem que não exista a variável no imóvel, se computava 1 (um), possível ausência da variável;
- caso não fosse possível determinar se a variável estava ou não presente, se computava 2 (dois), total imparcialidade;
- caso não houvesse certeza da existência da variável, mas as características do imóvel indicassem a existência da variável no imóvel, se computava 3 (três), possível presença da variável;

- caso a variável estivesse presente, se computava 4 (quatro), certeza da presença da variável.

### 3.2 Consolidação das variáveis

Durante o processo de regressão, constatou-se que, devido ao fato de diversos de seus dados serem notificados no site de forma errônea ou subnotificados, foi necessário excluir da análise as variáveis: Área total, IPTU, Ano de Construção e Churrasqueira. A variável churrasqueira foi excluída, pois se constatou a impossibilidade de distinção entre a existência de churrasqueira individual (dentro do apartamento) ou coletiva do condomínio.

Como várias variáveis não foram individualmente significativas para as análises iniciais de regressão, estas foram agrupadas visando ao melhor desempenho do modelo, segundo a sua associação aos conceitos abaixo:

a) Criada a variável INFRA (infraestrutura) formada pela adição das variáveis: Portaria24h, Elevador, Interfone e Lavanderia Externa.

b) Criada a variável LAZER formada pela adição das variáveis: Brinquedoteca, Quadra, E. Gourmet, Academia, Sauna, Piscina, Playground, Salão de Festas, Salão de Jogo.

c) Criada a Variável SOFISTIC (sofisticação) formada pela adição das variáveis: Mobiliado, Área de Serviço, Arm. Cozinha, Arm. Embutido, Ar Condicionado, Copa, Lavabo, Sacada, Hidromassagem, Depósito e Quarto de Empregada.

d) A variável denominada Fator SE (fator socioeconômico) foi adicionada também à amostra; esta foi obtida a partir de dados encontrados no site do Atlas Ambiental do Município de São Paulo. O Fator SE é uma variável relacionada à externalidade das regiões de São Paulo; foi criada utilizando-se os valores encontrados como “fator principal” nas tabelas encontradas no site do atlas ambiental. Segundo é descrito no site do Atlas Ambiental de São Paulo, “O Fator Principal de cada distrito foi obtido pela somatória da multiplicação dos pesos dos indicadores pelos seus valores padronizados no respectivo distrito, gerando uma classificação segundo seu perfil socioeconômico em ordem crescente do pior para o melhor padrão.”

(SÃO PAULO, 2016). Esses pesos e valores estão disponíveis para consulta no mesmo site.

A variável Fator SE é a única variável de externalidades presente nos cálculos deste trabalho.

## 4 Resultados

Os dados amostrais obtidos foram importados para o software IBM SPSS (IBM, 2010) e regredidos pelo método Backward Stepwise, que regride suas variáveis e elimina as que não têm significância para o modelo de regressão por meio de testes estatísticos.

### 4.1 Modelos de regressão

Após tratamento adequado dos dados, dois modelos de regressão linear múltipla foram calculados utilizando-se todas as variáveis existentes para regressão: um modelo para aluguel e outro para venda.

Durante a realização dos cálculos dos modelos, foi constatado que 12 imóveis para aluguel e a mesma quantidade para venda presentes na amostra eram outliers, estes então foram excluídos da amostra, resultando numa amostra de 576 dados amostrais, divididos igualmente entre aluguel e venda.

O modelo de regressão de aluguel se encontra abaixo, expresso pela Equação 4. Este modelo apresentou 6 variáveis independentes estatisticamente significativas, como mostram a Equação 4 e a Tabela 1 referente a este modelo, a qual foi retirada do software SPSS:

$$\begin{aligned} \text{Preço Aluguel} = & -970,902 * (N^{\circ} \text{deDorm.}) \\ & + 35,184 * (\text{AreaUtil}) + 538,898 * (\text{Vagas}) + \\ & 180,2396 * (\text{LAZER}) + 737,593 * (\text{SOFISTIC}) \\ & + 223,496 * (\text{FatorSE}) \end{aligned} \quad (4)$$

Este modelo apresentou um poder de explicação ( $R^2_{\text{ajustado}}$ ) de 91,9%. As variáveis que não se mostraram significativas para este modelo foram Condomínio e INFRA. A multicolinearidade presente no modelo de regressão foi avaliada pelo *Variance Inflation Factor* (VIF) e se mostrou elevada, embora ainda aceitável.

**Tabela 1 – Modelo de regressão linear múltipla calculado para aluguel**

model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Number of regressed sample data	Collinearity Statistics	
	B	Std. Error	Beta				Tolerance	VIF
Nº Dorm.	-970,902	107,449	-,518	-9,036	,000	288	,089	11,181
Area Util	35,184	2,022	,893	17,401	,000	288	,112	8,947
Vagas	538,898	136,384	,192	3,951	,000	288	,124	8,058
LAZER	180,239	80,123	,068	2,250	,025	288	,326	3,063
SOFISTIC	737,593	137,745	,288	5,355	,000	288	,101	9,856
Fator SE	223,496	72,790	,076	3,070	,002	288	,486	2,058

Fonte: Dados da pesquisa.

**Tabela 2 – Modelo de regressão linear múltipla calculado para venda**

model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Number of regressed sample data	Collinearity Statistics	
	B	Std. Error	Beta				Tolerance	VIF
Area Util	9637,616	356,172	,923	27,059	,000	288	,119	8,404
Vagas	170811,809	27175,310	,217	6,286	,000	288	,116	8,638
LAZER	60598,731	18580,299	,062	3,261	,001	288	,384	2,607
Fator SE	121135,744	14461,098	,125	8,377	,000	288	,621	1,611
Nº Dorm.	-174842,055	15752,104	-,314	-11,100	,000	288	,172	5,801

Fonte: Dados da pesquisa.

**Tabela 3 – Modelo de regressão linear múltipla da rentabilidade**

model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Number of regressed sample data	Collinearity Statistics	
	B	Std. Error	Beta				Tolerance	VIF
(Constant)	3,880E-3	6,0412E-5		64,219	,000			8,404
Nº Dorm	-,4,830E-4	2,589E-5	-,587	-18,656	,000	,587	1,705	8,638
Vagas	-5,827E-5	2,592E-5	-,079	-2,248	,025	,475	2,103	2,607
FatorSE	-3,514E-4	1,788E-5	-,627	-19,652	,000	,570	1,755	1,611
LAZER	-6,668E-5	1,926E-5	-,090	-3,461	,001	,860	1,163	
SOFISTIC	9,933E-4	3,740E-5	,895	26,555	,000	,511	1,955	5,801

Fonte: Dados da pesquisa.

Já o modelo de vendas apresentado na Equação 5 e na Tabela 2 apresentou 5 variáveis independentes estatisticamente significativas, a saber:

$$\begin{aligned} \text{Preço Venda} = & -174842,055*(N^\circ \text{ de Dorm.}) + \\ & 9637,616*(\text{AreaUtil}) + 170811,809*(\text{Vagas}) + \quad (5) \\ & 60598,731*(\text{LAZER}) + 121135,744*(\text{Fator SE}) \end{aligned}$$

O modelo de venda apresentou um poder de explicação ( $R^2_{\text{ajustado}}$ ) de 96,2%, e as variáveis que não se mostraram significativas foram SOFISTIC, INFRA e Condomínio.

É possível notar, ao analisar estes modelos, que estes possuem uma multicolinearidade elevada, porém aceitável. A constante de regressão atenuava a multicolinearidade nos dois modelos; apesar disso, um modelo cuja reta passa pela origem (modelo sem a constante de regressão) é interessante para a análise deste trabalho devido à premissa de que, se existir um hipotético imóvel “nulo” (nenhum dormitório, nenhuma área etc.), ou seja, que não tenha nenhuma dimensão, terá seu valor de mercado nulo. Um modelo que não tenha a constante de regressão se mostra mais interessante para este estudo, o que justifica, portanto, nossa opção.

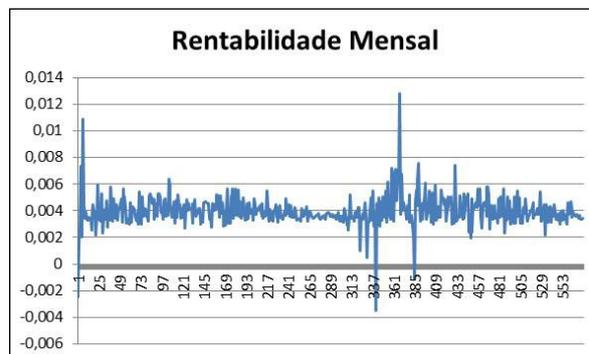
#### 4.2 Comparativo entre os modelos

Com os dois modelos já definidos, foram calculados os valores preditivos, tanto de aluguel como de venda, para todos os imóveis presentes na amostra, exceto os outliers, que ficaram definidos durante o tratamento dos dados. Utilizando estes valores obtidos, a razão aluguel sobre venda (Aluguel/Venda) foi calculada. A esta razão se pode denominar rentabilidade mensal do investimento imobiliário, a partir da premissa de que o aluguel atribuído ao imóvel é dividido por seu valor total, gerando, assim, uma taxa de recuperação mensal do investimento, ou rentabilidade mensal.

Esta rentabilidade mensal não é a rentabilidade real obtida no investimento imobiliário, pois não foram levados em consideração impostos e outros gastos envolvidos no processo, como valorização, depreciação etc. Esta rentabilidade (?), contudo, pode ser utilizada para o objetivo definido deste trabalho. Assim, esta abordagem pode ser entendida como um estudo exploratório de casos múltiplos com níveis analíticos imbricados.

Os valores da rentabilidade obtida são mostrados em um gráfico de dispersão, presente na Figura 2, a saber:

**Figura 2** – Comportamento da rentabilidade mensal para apartamentos residenciais



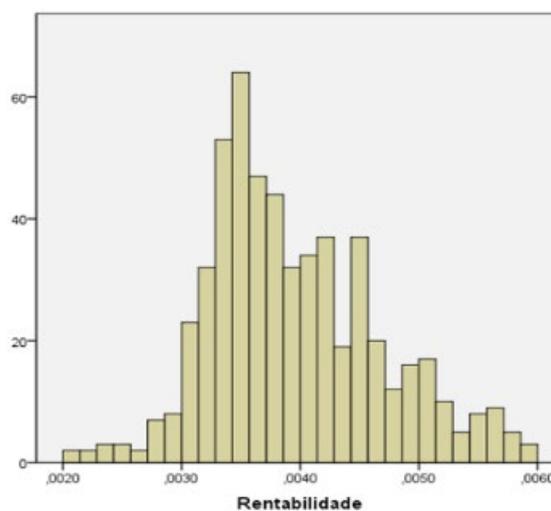
Fonte: Autoria própria

Percebe-se neste gráfico a presença de valores muito distintos do comportamento do restante da amostra; estes valores foram excluídos por apresentarem um comportamento não representativo do conjunto dos demais.

Percebe-se também, por este gráfico, que grande parte dos valores obtidos pela razão, ou rentabilidade mensal, se encontra no intervalo de 0,002 a 0,006. Os valores que estavam além deste intervalo são os que foram excluídos por se apresentarem dispersos e em casos isolados.

Com a exclusão de pontos atípicos e de valores que não se encontravam no intervalo definido, de 0,002 a 0,006, um novo gráfico foi construído, apresentado na Figura 3, abaixo, no formato de histograma:

**Figura 3** – Distribuição de frequência da rentabilidade mensal para apartamentos residenciais

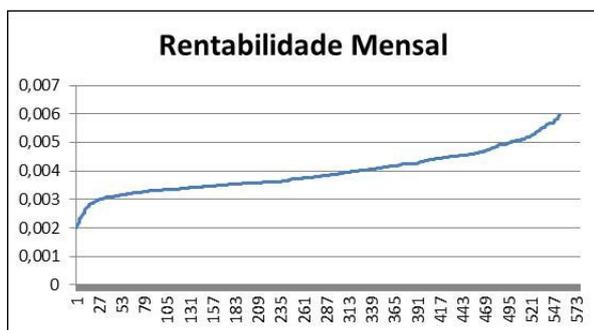


Fonte: Autoria própria

Ao analisar este histograma é possível visualizar que o intervalo definido para a rentabilidade, de 0,002 a 0,006, foi definido de forma correta. Esta afirmação é válida, já que, ao se analisarem as caudas deste histograma, percebe-se que estas são decrescentes e próximas de sua convergência.

Para efeitos de análise, os dados foram rearranjados de maneira que se pudesse verificar o comportamento do mercado imobiliário para apartamentos residenciais na cidade de São Paulo/SP. Estes dados foram expressos em um gráfico de dispersão, porém em ordem crescente para os valores da razão, conforme Figura 4 abaixo:

**Figura 4 –** Comportamento da rentabilidade mensal em ordem crescente para apartamentos residenciais



Fonte: Autoria própria

Estes gráficos expressam a rentabilidade referente a 92,5% dos 600 imóveis amostrados originalmente.

Ao analisar estes gráficos, juntamente com os dados amostrais, foi perceptível que o comportamento da rentabilidade mensal ideal no mercado imobiliário tende a ser menor em apartamentos mais bem providos de espaço e com maior valor de mercado, tanto para aluguel quanto para venda.

Analogamente, os dados mostraram que imóveis menos providos de espaço e com valor de mercado menor tendem a apresentar uma rentabilidade mensal ideal maior.

### 4.3 Validação dos modelos de regressão

Com a necessidade de validar os modelos de regressão de venda e aluguel, estes que são a base para todo o estudo feito, um reteste foi executado. Este reteste consistiu em coletar uma nova amostra de 150 imóveis – 75 imóveis anunciados para aluguel e 75 imóveis para venda. Entre estes 150 imóveis, foram escolhidos 25 de cada grupo, sendo estes os grupos V, III e grupo I + II, tanto para aluguel quanto para venda.

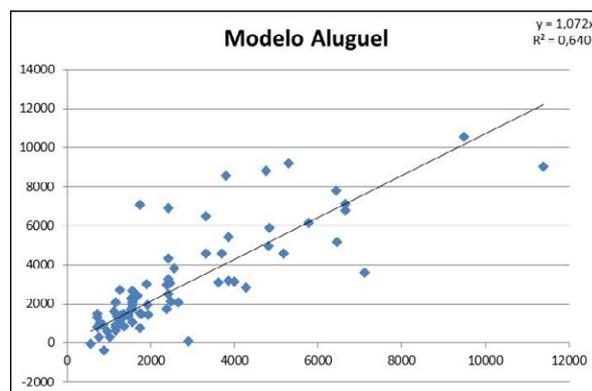
Estes imóveis foram coletados utilizando-se o mesmo processo da amostra original, ou seja, um gerador de números aleatórios para a definição dos imóveis coletados. Com o intuito de ser o mais abrangente possível, os imóveis coletados para o reteste foram de distintas regiões de São Paulo, excluindo os bairros já presentes na amostra original, à exceção do grupo I+II.

Devido a fatores como a valorização/desvalorização de mercado, os valores de aluguel e venda tiveram de ser convertidos para a mesma época da coleta da amostra original, para que não houvesse erros de previsão por conta deste fato. Para isso, utilizou-se o índice IGP-M para converter o valor de aluguel, e, para vendas, o INCC, dividindo-se o valor do imóvel pelo índice obtido.

A escolha destes índices foi baseada em pesquisas dos índices mais usuais no município de São Paulo, entretanto, não há um índice usual de preço de vendas de imóveis já construídos. Por este motivo, foi definida a utilização do INCC, pois este acompanha a inflação do mercado.

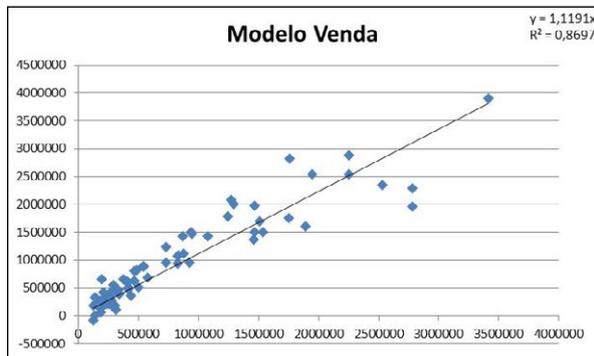
Com a nova amostra coletada, foram calculados os valores preditivos, utilizando-se os modelos de regressão de venda e aluguel. Com o intuito de verificar os resultados obtidos, foram construídos gráficos de dispersão dos valores coletados pelos valores preditivos, para aluguel e venda, presentes abaixo nas Figuras 5 e 6. Nestes gráficos foram traçadas retas médias para analisar (ou apresentar ?) os dados obtidos.

**Figura 5 –** Dispersão dos dados de aluguel



Fonte: Autoria própria

Figura 6 – Dispersão dos dados de venda



Fonte: Autoria própria

O poder de explicação da reta de aluguel foi de 64,01%, um resultado aceitável; já o poder de explicação da reta de venda foi de 86,97%, um ótimo resultado. Acredita-se que o motivo pelo qual a reta de aluguel explica menos que a de venda deva-se ao fato de que alguns fatores muito relevantes para a determinação do valor do aluguel, tais como o estado de conservação do imóvel e do condomínio, não puderam ser captados pelo meio utilizado para a coleta dos dados deste trabalho.

As duas retas mostram que as previsões para os valores no reteste são um pouco maiores que os valores coletados – em torno de 7,2% para aluguel e 11,9% para venda. Tais diferenças estão relacionadas às grandes oscilações da economia do Brasil, e, especificamente no mercado imobiliário, estas diferenças não são totalmente compensadas pelos indicadores oficiais IGP-M e INCC. Os resultados obtidos fornecem robustez à metodologia empregada neste trabalho, indicando coerência nos conceitos e nas premissas utilizadas para a obtenção do modelo de rentabilidade.

#### 4.4 Análise de rentabilidade no mercado imobiliário

Para entender quais variáveis explicavam a rentabilidade existente na locação de apartamentos residenciais de São Paulo/SP, um novo modelo hedônico foi calculado, este tendo a rentabilidade obtida através da razão dos outros dois modelos, como variável dependente, e as demais variáveis já citadas como variáveis independentes.

Este modelo, descrito na Equação 6, apresentou 5 variáveis independentes e uma constante de regressão estatisticamente significativas, a saber:

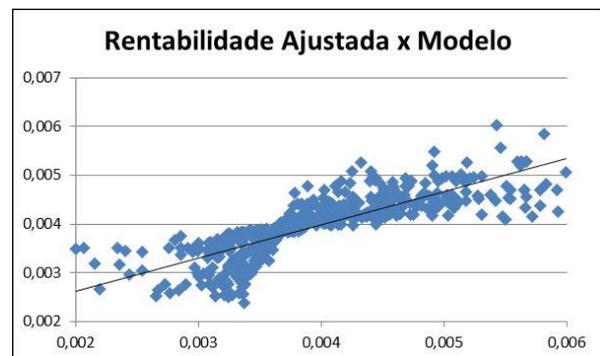
$$\begin{aligned} \text{Rentabilidade} = & [388 - 48,30*(N^\circ \text{deDorm.}) \\ & - 5,827*(Vagas) - 35,14*(Fator \text{ SE}) \\ & - 6,668*(LAZER) + 99,33*(SOFIS \text{ TIC})] \end{aligned} \quad (6)$$

Este modelo apresentou um poder de explicação ( $R^2_{ajustado}$ ) de 67,9%, um resultado aceitável, tendo em consideração que este modelo foi obtido a partir de valores preditos a partir de outros modelos. As variáveis que não se mostraram significativas para este modelo foram Área Útil, Condomínio e INFRA. A Tabela 3 apresenta o modelo da rentabilidade calculado.

A constante de regressão foi mantida neste modelo, uma vez que os valores de rentabilidade obtidos na razão entre os modelos de Aluguel/Venda variavam no intervalo definido, de 0,2% a 0,6%, nunca assumindo um valor nulo. Segundo este raciocínio, o fato de não poder existir um valor nulo para a rentabilidade está atribuído ao fato de esta ser uma divisão (aluguel/venda). Assim sendo, se existisse um imóvel nulo, esta divisão seria indeterminada, entretanto o limite desta razão para este imóvel hipotético convergiria para um valor real não nulo. Assim, por conta destes fatos, é mandatório manter a constante de regressão.

Na Figura 7, a seguir, é possível observar este poder de explicação quando plotado um gráfico de dispersão dos valores preditos, obtidos com modelo de rentabilidade *versus* a rentabilidade calculada através dos valores preditos dos modelos de aluguel e venda.

Figura 7 – Rentabilidade prevista pelo modelo x Rentabilidade obtida pela razão dos valores preditos pelos modelos Aluguel/Venda



Fonte: Autoria própria

Pelas características inerentes aos modelos de regressão, a razão entre duas retas regredidas resulta em uma distribuição não homogênea dos resíduos,

sendo observada uma dispersão tanto maior quanto mais distante dos valores médios esperados.

#### 4.5 Características para maior rentabilidade

A partir do coeficiente beta reduzido do modelo de regressão da rentabilidade, presente na Tabela 3, de regressão, foi possível determinar quais variáveis são mais relevantes para uma melhor rentabilidade em investimentos imobiliários. Este coeficiente fornece a importância relativa de cada uma das variáveis independentes para o modelo de regressão explicativo da variável dependente.

De acordo com os resultados deste modelo, as características influenciadoras na rentabilidade de investimentos imobiliários, em módulo e em ordem decrescente, são: maior sofisticação, menor fator socioeconômico, menor número de dormitórios, menor quantidade de lazer e menor número de vagas.

A variável relacionada à sofisticação apresenta o maior coeficiente beta reduzido e este positivo. Tal fato se relaciona à qualidade do acabamento do imóvel bem como ao fato de este estar ou não mobiliado. As demais variáveis apresentam um coeficiente beta reduzido negativo.

Assim, as características imobiliárias que oferecem maior rentabilidade são:

- Maior sofisticação: quanto mais benfeitorias e itens de conforto do imóvel, maior será seu valor de aluguel em relação ao seu valor de venda. Isto ocorre porque as benfeitorias nos imóveis são de valor muito menor quando comparadas com o valor de venda do imóvel, e estas possuem um grande peso no valor de uso atribuído na locação, fazendo com que investir nessas benfeitorias acresça o valor do aluguel significativamente, melhorando a rentabilidade.
- Fator socioeconômico mais baixo: imóveis em lugares onde o fator socioeconômico é mais baixo possuem uma rentabilidade maior. Este fato está relacionado à proporção entre o aluguel e a venda em distintas regiões de São Paulo (local estudado), de forma que, numa região onde o fator socioeconômico é mais baixo, o capital investido é menor e a recuperação deste capital é maior quando comparado com um investimento em imóvel localizado em uma região de fator socioeconômico mais alto.

- Menor número de dormitórios: quanto menor o número de dormitórios no imóvel, maior será a rentabilidade. Isto se dá porque imóveis com maior número de dormitórios necessitam de mais área útil, encarecendo o valor de venda do imóvel, não sendo compensado no valor do aluguel proporcionalmente, gerando assim uma menor rentabilidade por número de dormitórios.

- Menos lazer: imóveis com menos lazer possuem maior rentabilidade. Este fato está relacionado ao custo-benefício no investimento. Imóveis que possuem mais lazer tendem a se localizar em condomínios mais sofisticados e, portanto, com valores de venda e de condomínio encarecidos.

- Menor quantidade de vagas de garagem: quanto menor o número de vagas, maior será a rentabilidade. Isto ocorre porque, quanto maior a quantidade de vagas, maior será o valor de venda e de locação do imóvel; porém este aumento possui uma disparidade, sendo a rentabilidade associada à vaga de garagem menor que a do apartamento em si, portanto o incremento no número de vagas diminui a rentabilidade no investimento.

## 5 Conclusão

Após as análises dos modelos de regressão linear múltipla obtidos, foi possível determinar as características hedônicas imobiliárias que trazem maior rentabilidade para o investimento em apartamentos no município de São Paulo.

Estas características são, em ordem decrescente de significância: maior sofisticação no imóvel, menor fator socioeconômico, menos dormitórios no imóvel, menor quantidade de lazer no condomínio e menos vagas de garagem. Estas características em um imóvel geram maior rentabilidade no investimento imobiliário, segundo os cálculos efetuados.

Imóveis destinados à locação no município de São Paulo, visando a rendimento, devem ser preferencialmente, e nesta ordem: bem cuidados e equipados; em regiões de perfil socioeconômico modesto; com poucos dormitórios; em condomínios relativamente simples; e com poucas ou nenhuma vaga de garagem.

Estes resultados sugerem a necessidade de estudos mais abrangentes referentes às causas dessa variação de rentabilidade, tais como a associação

entre a rentabilidade e o risco de inadimplência ou a expectativa de valorização futura do imóvel.

Tais resultados mostraram que o método apresentado neste trabalho é consistente e estes modelos e resultados obtidos podem ser aplicados para distintas regiões de São Paulo/SP.

## REFERÊNCIAS

ABRAMO, P.; FARIA, T. C. Mobilidade residencial na cidade do Rio de Janeiro: considerações sobre os setores formal e informal do mercado imobiliário. *In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS DA ABEP*, 11., 1998, Caxambu, MG. **Anais [...]** Belo Horizonte: ABEP, 1998. p. 421-456.

ALVES, V. **Avaliação de imóveis urbanos baseada em métodos estatísticos multivariados**. 2005. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, Campo Mourão, 2005.

ANGELO, C. F.; FÁVERO, L. P. L. Modelo de preços hedônicos para a avaliação de veículos novos. *In: VI SEMINÁRIO EM ADMINISTRAÇÃO FEA-USP*, 2003, São Paulo. **Relação de trabalhos**, São Paulo: VI SEMEAD, 2003. 1 CD-ROM.

BAPTISTELLA, M. **O uso de redes neurais e regressão linear múltipla na engenharia de avaliações**: determinação dos valores venais de imóveis urbanos. 2005. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, Curitiba, 2005.

BRAULIO, S. N. **Proposta de uma metodologia para a avaliação de imóveis urbanos baseado em métodos estatísticos multivariados**. 2005. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, Curitiba, 2005.

BRONDINO, N. C. M.; SILVA, A. N. R. Da teoria à prática: introduzindo técnicas emergentes na avaliação de imóveis em massa. **Revista Minerva – Pesquisa & Tecnologia**, v. 2, n. 2, p. 107-117, 2005.

CARMO, C. R. S. Precificação imobiliária baseada em modelagem hedônica e externalidades: um estudo aplicado a terrenos urbanos. **Revista da Faculdade de Administração e Economia**, v. 5, n. 2, p. 2-23, 2014.

DANTAS, R. A. **Engenharia de Avaliações**: uma introdução à metodologia científica. São Paulo: Pini, 1998.

FÁVERO, L. P. L. **Modelos de preços hedônicos aplicados a imóveis residenciais em lançamento no município de São Paulo**. 2003. Dissertação (Mestrado em Administração) – Universidade de São Paulo, São Paulo, 2003.

GAZOLA, S. (2002). **Construção de um modelo de regressão para avaliação de imóveis**. 2002. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2002.

GUERRA, M. J.; DONAIRE, D. **Estatística indutiva**: teoria e aplicações. 5. ed. São Paulo: Livraria Ciência e Tecnologia Editora, 1991. v. 2.

MARQUES, J. L.; CASTRO, E. A.; BHATTACHARJEE, A. A localização urbana na valorização residencial: Modelos de autocorrelação espacial. *In: CONGRESSO DA ASSOCIAÇÃO PORTUGUESA DE DESENVOLVIMENTO REGIONAL*, 15., 2009, Cidade da Praia, Cabo Verde. **Actas [...]** Cidade da Praia: APDR, 2009. p. 2224-2244.

MOREIRA, D. S.; SILVA, R. S.; FERNANDES, A. M. R. Engenharia de avaliações de imóveis apoiada em técnicas de análise multicritério e redes neurais artificiais. **Revista de Sistemas de Informação da FSMA**, n. 6, p. 49-58, 2010.

NEGRI NETO, A. Preços hedônicos. *Informações Econômicas*, v. 33, n. 12, p. 81-83, dez. 2003.

ROSEN, S. Hedonic prices and implicit markets: product differentiation in pure competition. **Journal of Political Economy**, v. 82, n. 1, p. 34-55, 1974.

SANTELLA, R. **Avaliação de imóveis urbanos com utilização da lógica difusa**. 2004. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

SÃO PAULO (Município). **Atlas Ambiental do Município de São Paulo**. 2016.

Disponível em: <http://atlasambiental.prefeitura.sp.gov.br/>. Acesso em: 23 fev. 2016.

IBM. IBM SPSS Statistics. Versão 19. Armonk, NY: IBM, 2010.

ZAP. **ZAP Imóveis**. Disponível em: <http://www.zapimoveis.com.br/>. Acesso em: 21 out. 2015.