



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Centro de Ciências Jurídicas e Econômicas
Faculdade de Administração e Ciências Contábeis - FACC

João Lucas Gemmal Pinto

SEGMENTAÇÃO E AS DINÂMICAS REGIONAIS DO MERCADO
IMOBILIÁRIO CARIOCA

Rio de Janeiro – RJ

2023

JOÃO LUCAS GEMMAL PINTO

**SEGMENTAÇÃO E AS DINÂMICAS REGIONAIS DO MERCADO
IMOBILIÁRIO CARIOCA**

Monografia apresentada como requisito parcial à obtenção do grau de Bacharel em Administração à Faculdade de Administração e Ciências Contábeis da Universidade Federal do Rio de Janeiro (FACC/UFRJ).

Orientadora: Cristina Pimenta de Mello Spineti Luz

Rio de Janeiro - RJ

2023

RESUMO

O mercado imobiliário é tão complexo quanto fundamental a vida de todo cidadão. O acesso a um imóvel influencia também o acesso do cidadão a saúde, lazer e cultura. Para auxiliar a compreensão das dinâmicas deste mercado na cidade do Rio de Janeiro, o presente trabalho utiliza de métodos de regressão linear múltipla aplicada no estudo do preço de imóveis à venda no mês de outubro de 2022. Os dados coletados foram tratados com o intuito de se obter o modelo que melhor represente os imóveis em estudo. Foram testados três modelos para representar as zonas do município, por região, distrito e por bairro, e ao todo em torno de 150 variáveis foram testadas, quinze das quais foram selecionadas para a equação. O modelo utilizado foi construído através da regressão linear múltipla. Os resultados demonstraram que a localização é o maior determinante no preço, enquanto características intrínsecas tem uma influência marginal.

Palavras-chave: regressão linear multipla; avaliação de imóveis; mercado imobiliário.

LISTA DE QUADROS

Quadro 1 - Referencial Teórico..... 13

LISTA DE TABELAS

Tabela 1 - Distribuição dos domicílios por condição de ocupação no município do Rio de Janeiro.....	21
Tabela 2 - Preço por metro quadrado no Rio de Janeiro por bairros selecionados...	21
Tabela 3 - Dez Bairros mais representados pela amostragem.....	25
Tabela 4 - Estatística Descritiva do recorte final dos dados	26
Tabela 5 – Resultados regressão do recorte de região por tipo de imóvel	31
Tabela 6 - Resultados por recorte geográfico	32
Tabela 7 - Variáveis selecionadas pela seleção progressiva	36
Tabela 8 - Variáveis selecionadas pela seleção regressiva	37
Tabela 9 - Coeficiente Estimados.....	38

LISTA DE FIGURAS

Figura 1 - Mapa do município do Rio de Janeiro	20
Figura 2 - Coeficiente de correlação das variáveis quantitativas do modelo	34

LISTA DE GRÁFICOS

Gráfico 1 - Boxplot dos preços por Região.....	27
Gráfico 2 - Boxplot dos preços por Tipo	28
Gráfico 3 – Dispersão dos preços por Metro quadrado	29
Gráfico 4 - Resíduos por preço estimado	33
Gráfico 5 - Resíduos por preço estimado após transformação	33
Gráfico 6 - Resíduo por Área.....	35

SUMÁRIO

1. INTRODUÇÃO	9
1.2. Objetivos	11
1.2.1. Objetivo Geral	11
1.2.2. Objetivos Específicos	11
1.3 Justificativas.....	11
2. FUNDAMENTAÇÃO TEÓRICA	13
2.1. Avaliação de imóveis	15
2.2. Regressão Linear Múltipla.....	16
2.2.1.1 Linearidade	17
2.2.1.2 Multicolinearidade	18
2.2.1.3 Homocedasticidade.....	19
2.2.2 Modelo de Preços Hedônicos	19
2.3. Visão Geral do Mercado Imobiliário da Cidade do Rio de Janeiro	19
3. METODOLOGIA	22
3.1. Método, Classificação e Técnicas de pesquisa.....	22
3.2. Coleta dos Dados	22
3.3. Seleção das Variáveis	23
3.4. Análise dos dados.....	23
4. RESULTADOS.....	25
4.1. Estatística Descritiva	25
4.2. Regressão Linear Múltipla.....	29
4.2.1 Equação do Modelo	29
4.2.2 Resultados	30
4.2.3 Verificação dos Pressupostos	32
4.2.3.1 Homocedasticidade.....	32

4.2.3.2 Colinearidade.....	34
4.3. Seleção de Variáveis	35
4.4 Considerações Finais	37
5. CONCLUSÃO	40
REFERÊNCIAS.....	41
APÊNDICE A - Tabela com todos os coeficientes de regressão da Equação 4	
413	

1. INTRODUÇÃO

O setor imobiliário tem grande influência na vida de todos os cidadãos e na organização urbana. Das suas características destacam-se o grande investimento que a aquisição de um imóvel representa para o seu consumidor médio ou de renda, no caso da locação, e a complexidade de fatores que podem influenciar a formação de seu preço. Quando o consumidor opta por um imóvel ele tem de considerar tanto as características intrínsecas àquela unidade, quanto ao ambiente ao redor dela, as amenidades e externalidades negativas daquela região.

A formação de preço influencia a formação social dos habitantes de cada região, pois vai determinar quem terá e quem não terá acesso à moradia na região, levando à criação de padrões na ocupação do solo. Em uma metrópole como o Rio de Janeiro, onde populações de diversas origens e rendas se encontram, cria-se um ambiente propício à segregação desses espaços, como pontua Alves et al. (2011) em seu estudo sobre a cidade de São Paulo.

De acordo com o último censo do IBGE (2010), a cidade do Rio de Janeiro abriga em torno de seis milhões de habitantes e mais de dois milhões de domicílios. Por causa dessa concentração populacional e do seu status como polo cultural e turístico, a cidade contém o segundo metro quadrado mais caro do Brasil, de acordo com o índice FIPEZAP (FIPE, 2022). Também contém no município o bairro mais caro a se morar em todo o Brasil: o Leblon (FIPE, 2022).

Contudo, o desempenho da economia e do mercado imobiliário da cidade foram muito afetados pela crise econômica de 2017 e a pandemia do coronavírus. Ao observar dados do índice FIPEZAP (FIPE, 2022) é possível dividir a cronologia do mercado imobiliário brasileiro no século XXI em três momentos. Primeiro, o mercado imobiliário brasileiro passou por um período de crescimento vertiginoso entre 2003 e 2015. Isso ocorreu devido a diversos fatores, como mudanças no mercado de crédito habitacional, queda na taxa de juros e aumento da renda, que somados aumentaram muito a demanda por bens imobiliários. O fim desse período de crescimento no mercado coincide com o começo da crise econômica. E o terceiro período é marcado por uma retomada lenta.

No período entre 2015 e 2021, dentre as capitais brasileiras, o Rio de Janeiro destaca-se como tendo alcançado um decréscimo no preço imobiliário durante cinco anos, de 2015 a 2020 (FIPE, 2022). Este fenômeno pode ser explicado pelo

maior impacto da crise no Estado do Rio de Janeiro devido a sua economia ser dependente da indústria de extração do petróleo, desgaste político e falta de confiança dos consumidores e investidores no Estado. Para Brando e Barbedo (2016), a variação dos preços dos imóveis no tempo pode ser explicada tanto por variáveis econômicas, sendo as mais comuns a oferta de imóveis e a renda das famílias, quanto por variáveis comportamentais, como a expectativa e como a região é vista pelos consumidores.

A partir de 2021, ocorre uma retomada lenta da economia do município. São divulgados novos lançamentos imobiliários e os preços do mercado passam por variação positiva tímida, ainda abaixo da inflação.

O estudo sobre o comportamento dos preços de imóveis é de grande importância para a economia, pois é através destes estudos que se torna possível fazer estimações sobre a demanda do setor e preferências dos consumidores de uma região. O preço de um imóvel pode ser formado por diversos fatores, como a situação econômica da região, o conjunto de atributos físicos daquele bem e as preferências dos consumidores. Excluindo os atributos físicos, todos os fatores podem variar entre regiões e até em longos intervalos de tempo, o que gera incerteza nos investidores.

Entre os métodos para estudar o comportamento dos preços dos imóveis, o mais comum é a regressão linear múltipla, segundo a Associação Brasileira de Normas Técnicas (ABNT, 2011).

Na literatura também é possível encontrar trabalhos que se utilizam desta regressão para estudar os preços de venda e locação, a partir dos modelos de preços hedônicos, visando encontrar o preço das características individuais (PINTO e FERNANDES, 2019; ROSA; OLIVEIRA; PINTO; 2019; RIBEIRO, LUPORINI, 2019). A vantagem deste método está em valorar as características individuais, entender quais atributos são mais valorizados no mercado e a disposição do consumidor em pagar por um incremento entre moradias. Outros métodos como a regressão quantílica, utilização de redes neurais artificiais e o método de vendas repetidas também são utilizados em estudos da área.

Diante desse contexto, este trabalho pretende aumentar a compreensão sobre a formação de preços dos imóveis no Rio de Janeiro e buscar padrões nas ofertas de imóveis a partir de classificados online, que são cada vez mais prevalentes nesse mercado.

Assim, o presente estudo aplica modelos estatísticos de regressão para explicar os preços dos imóveis à venda na cidade do Rio de Janeiro e a precificação dos imóveis nas diferentes sub-regiões da cidade. A aplicação destes modelos visa expandir a compreensão sobre o mercado imobiliário e as variáveis que influenciam o preço dos imóveis na cidade. Devido à grande diversidade e desigualdade socioeconômica da cidade torna-se essencial a segmentação desta em sub-regiões, para avaliar a existência ou não de resultados distintos para as mesmas variáveis. Portanto, a pergunta que este trabalho busca responder é: como pode ser entendida a precificação de imóveis nos diferentes mercados imobiliários cariocas, a partir de modelos de regressão?

1.2. Objetivos

Esta seção é dedicada a descrever os objetivos da pesquisa.

1.2.1. Objetivo Geral

O objetivo geral desta pesquisa é mapear diferentes variáveis que afetam o preço de venda de imóveis residenciais na cidade do Rio de Janeiro.

1.2.2. Objetivos Específicos

Os objetivos específicos desta pesquisa são:

- Segmentar a cidade em diferentes sub-regiões ou zonas de valor;
- Identificar os bairros mais valorizados de suas respectivas regiões.

1.3 Justificativas

Existem hoje dois índices de preços de imóveis estabelecidos no Brasil: o divulgado pela Fundação Instituto de Pesquisa Econômica (FIPE) e o Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R). Os dois índices utilizam do método da mediana estratificada para determinar os preços. Porém, a literatura acadêmica (DIEWERT, 2009, apud PAIXÃO; RIBEIRO; LUPORINI, 2019)

aponta que a utilização de métodos econométricos consegue gerar análises mais profundas. Dos métodos econométricos se destaca a regressão linear múltipla utilizada por diversos trabalhos nacionais como Paixão, Ribeiro e Luporini (2019), Pintos e Fernandes (2019), Malaman e Amorim (2017), entre outros.

Na literatura, os estudos brasileiros que utilizam regressão tendem a focar seu escopo em uma cidade e com uma amostra limitada, seja em tamanho ou por representar um momento anterior à crise que começou em 2015. O presente trabalho busca preencher essa lacuna ao utilizar dados atualizados para o período de 2022.

De acordo com Pinto e Fernandes (2019), o estudo do mercado imobiliário é relevante não apenas para os vendedores e compradores de imóveis e as construtoras, pois diversos órgãos governamentais usam do preço de mercado para estipular alienações, cobranças de impostos e taxas. O preço utilizado é o chamado valor venal, que seria o valor provável a ser praticado para a venda daquele imóvel no mercado e é calculado com base nos preços praticados no mercado imobiliário.

Portanto, é importante o estudo e aplicação de modelos estatísticos à realidade brasileira para que possa haver uma compreensão mais profunda do mercado, que sirva de base para a tomada de decisão tanto nas esferas cível, quanto governamental. Podendo, ainda, ser útil para investidores do setor dirimirem suas incertezas e consumidores compreenderem a formação dos preços.

2. FUNDAMENTAÇÃO TEÓRICA

Para o embasamento teórico deste trabalho foram utilizadas a busca do portal de periódicos CAPES e a biblioteca eletrônica Scielo, procurando-se pelas palavras-chaves: “regressão linear” em conjunção com “mercado imobiliário” ou “preços de imóveis” em conjunção com “regressão linear”, no título ou no resumo do artigo e selecionados os mais relevantes ao tema deste trabalho e que foram publicados depois do ano 2000. Depois da pesquisa inicial também foram procurados artigos que tivessem a expressão “preços hedônicos” em conjunto com “preços de imóveis” ou “mercado imobiliário”. Dessa primeira pesquisa foram encontrados oito artigos relevantes, desses oito, seis foram selecionados após leitura, além de referências destes (cruzada). Todos os artigos selecionados foram revisados por pares.

Quadro 1 - Referencial Teórico

SOBRENOME DOS AUTORES	ANO	PERIÓDICO	TÍTULO DA OBRA	BIBLIOTECA ELETRÔNICA	CLASSIFICAÇÃO DO PERIÓDICO
Rosa, Oliveira e Pinto	2019	REVISTA GEPROS	Modelos de precificação para locação e venda de imóveis residenciais na cidade de João Monlevade-MG via regressão linear	CAPES	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B3
Ribeiro e Luporini	2019	Nova Economia	A valorização imobiliária em Belo Horizonte, 1995-2012: uma análise hedônica-quantílica	SCIELO	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B1 ECONOMIA - B1
Pinto e Fernandes	2019	interações	Análise de preços hedônicos no mercado imobiliário residencial de Conselheiro Lafaiete, MG	SCIELO	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO -B1
Nunes, Neto e Freitas	2019	Ambiente Construído	Modelo de regressão linear múltipla para avaliação do valor de mercado de apartamentos residenciais em Fortaleza, CE	SCIELO	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B2

SOBRENOME DOS AUTORES	ANO	PERIÓDICO	TÍTULO DA OBRA	BIBLIOTECA ELETRÔNICA	CLASSIFICAÇÃO DO PERIÓDICO
Nunes, Neto e Freitas	2019	Ambiente Construído	Modelo de regressão linear múltipla para avaliação do valor de mercado de apartamentos residenciais em Fortaleza, CE	SCIELO	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B2
Medeiros e Carvalho	2017	Revista de Economia da UEG	Modelagem econométrica do preço de aluguéis de apartamentos na cidade de Petrópolis-RJ utilizando regressão linear múltipla	CRUZADA	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO B4
Stertz <i>et al.</i>	2016	GESTÃO.ORG	Mercado Imobiliário: uma Análise sobre o Comportamento dos Preços dos Imóveis na Cidade de Porto Alegre/RS	CAPES	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B2
Alves et al .	2011	Revista Brasileira de Finanças	Modelagem dos preços de imóveis residenciais paulistanos	CRUZADA	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B1
Arraes e Filho	2008	Economia Aplicada	Externalidades e formação de preços no mercado imobiliário urbano brasileiro: um estudo de caso	CRUZADA	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - B1
Lacerda e Funcia	2005	Cadernos EBAPE.BR	O impacto da violência criminal urbana no preço dos imóveis residenciais na região da Tijuca, cidade do Rio de Janeiro: um estudo exploratório	SCIELO	ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO - A2

Fonte: elaborado pelo autor (2023).

2.1. Avaliação de imóveis

É importante destacar a peculiaridade dos imóveis como bem comercializado. Por ser um bem durável, o imóvel é muito comparado com outros bens duráveis como os automóveis, porém o imóvel incorre em algumas características únicas (PINTO; FERNANDES, 2019). Uma delas é sua dupla funcionalidade, tanto como um bem de consumo, quanto um investimento ou reserva de valor. Outra peculiaridade é o fato de ser impossível encontrar dois imóveis idênticos, o que tornaria difícil a comparação. Imóveis podem ter várias características iguais, como dois apartamentos no mesmo condomínio terem a mesma área, número de quartos, banheiros e vagas de estacionamento, mas ainda assim outras características permitem a diferenciação, como: em que andar se encontra o apartamento, o estado em que ele se encontra, para qual lado do edifício apontam suas janelas. Por causa disso, alguns autores como Ribeiro e Luporini (2019) classificam-nos como bens diferenciados.

No tocante às características que influenciam no preço dos imóveis, a literatura (LACERDA e FUNCIA, 2005) aponta para quatro categorias de características: estruturais, locacionais, de vizinhança e ambientais. Estruturais são aqueles relacionados às características individuais daquela unidade, locacionais à qualidade do acesso ao transporte e distância dos centros comerciais, da vizinhança é o acesso a escolas de qualidade, saúde e segurança pública e, por último, o ambiental é o nível de poluição e a vista da região (LACERDA e FUNCIA, 2005; ROSA; OLIVEIRA; PINTO, 2019).

Trabalhos como o de Rosa, Oliveira e Pinto (2019) e de Pinto e Fernandes (2019) utilizaram como variáveis o número de quartos e vagas na garagem, ocorrências policiais, proximidade com o centro comercial, postos de saúde e escolas estaduais mais próximas, e apontam que as características físicas tem uma maior influência na formação dos preços.

Lacerda e Funcia (2005) no seu artigo exploratório procuraram estudar a relação entre segurança e a variação de preços e encontraram uma relação entre maior número de crimes e diminuição no valor de compra de imóveis. No mesmo estudo foi utilizado o CEP como variável para assim melhor mapear as diferenças socioeconômicas.

Nunes, Neto e Farias (2019) apresentam um trabalho com uma amostra específica de apenas lançamentos de apartamentos na cidade de Recife, com esse

recorte foram capazes de medir variáveis restritas à categoria como número de torres, apartamentos por andar, número de andares e total de unidades de um mesmo tipo de apartamento. Os autores também utilizaram duas variáveis para representar a segmentação no mercado imobiliário, uma para representar a segmentação econômica, dividindo as unidades em categorias relacionadas aos seus preços. A outra visou uma segmentação espacial por variáveis *dummies* que representaram as diferentes zonas de valor utilizadas pelas imobiliárias da cidade.

Em Ribeiro e Luporini (2019) argumenta-se que o mercado imobiliário de uma dada cidade é composto de diferentes submercados. Os submercados são vizinhanças cujos os imóveis compartilham das mesmas características locacionais e podem ser considerados substitutos entre si. Esta dinâmica aponta para uma estrutura urbana segregada. Esta segregação é respaldada também pelos resultados da pesquisa de Lacerda e Funcia (2005) que apontam para uma diminuição dos preços dos imóveis quando estes se localizam próximos de favelas.

2.2. Regressão Linear Múltipla

A regressão linear múltipla é o método mais utilizado com o objetivo de estudar o comportamento de uma variável dependente preço em relação a outras variáveis (ABNT, 2011). As etapas para a construção de um modelo de regressão para o estudo do mercado imobiliário foram definidas por Gazola (2002) como: identificar as variáveis independentes, fazer o levantamento de dados, transformar as variáveis, fazer uma análise exploratória, construir o modelo de regressão, realizar a análise crítica das variáveis, analisar os resíduos e verificar a aplicabilidade do modelo.

2.2.1 Pressupostos

O modelo de regressão linear deve seguir alguns preceitos básicos para garantir que seus resultados sejam confiáveis e consistentes. Em suma, deve-se observar para que haja uma relação linear entre a variável dependente e as explicativas, os erros sejam independentes e tenham uma distribuição normal, com média zero e variância constante (homocedasticidade), não ocorra multicolinearidade ou autocorrelação entre as variáveis explicativas. Esses pressupostos são definidos pela ABNT, na norma NBR 14653-2, como:

"a) para evitar a micronumerosidade, o número mínimo de dados efetivamente utilizados (n) no modelo deve obedecer aos seguintes critérios, com respeito ao número de variáveis independentes (k):

$n \geq 3(k + 1)$ para $n \leq 30$,

$n_i \geq 3$ para $30 < n \leq 100$,

$n_i \geq 10\% n$ para $n > 100$, $n_i \geq 10$

onde n_i é o número de dados de mesma característica, no caso de utilização de variáveis dicotômicas e variáveis qualitativas expressas por códigos alocados ou códigos ajustados;

b) atentar para o equilíbrio da amostra, com dados bem distribuídos para cada variável no intervalo amostral;

c) os erros são variáveis aleatórias com variância constante, ou seja, são homocedásticos;

d) os erros são variáveis aleatórias com distribuição normal;

e) os erros são não autocorrelacionados, isto é, são independentes sob a condição de normalidade;

f) o engenheiro de avaliações deve se empenhar para que as variáveis importantes estejam incorporadas no modelo – inclusive as decorrentes de interação – e as variáveis irrelevantes não estejam presentes;

g) em caso de correlação linear elevada entre quaisquer subconjuntos de variáveis independentes, isto é, multicolinearidade, deve-se examinar a coerência das características do imóvel avaliando com a estrutura de multicolinearidade inferida, vedada a utilização do modelo em caso de incoerência;

h) não devem poder correlações evidentes entre o erro aleatório e as variáveis independentes do modelo, ou seja, o gráfico de resíduos não pode sugerir evidências de regularidade estatística com respeito às variáveis independentes;

i) possíveis pontos influenciadores, ou aglomerados deles, devem ser investigados e sua retirada fica condicionada à apresentação de justificativas." (ABNT, 2011, p. 34)

2.2.1.1 Linearidade

A linearidade significa a existência de uma relação linear entre a variável dependente e cada variável explicativa do modelo. Isto pode ser verificado de três maneiras (GAZOLA, 2002). Pela análise de gráficos de dispersão entre cada variável explicativa e a dependente, confirmando que ocorra um alinhamento entre os pontos no gráfico. A análise gráfica da dispersão do resíduo e cada variável independente com o objetivo de confirmar que o resíduo tem uma distribuição aleatória. Por fim, podemos utilizar do cálculo do coeficiente de correlação entre a variável dependente e as explicativas.

Quando é identificada a inexistência de linearidade, pode ser utilizada a transformação das variáveis explicativas a fim de linearizar o modelo. Gazola (2002) recomenda as seguintes transformações para uma dada variável independente X :

- Logarítmica, $X' = \log X$;

- Raiz quadrada, $X' = \sqrt{X}$;
- Quadrática, $X' = X^2$;
- Exponencial, $X' = \exp(X)$;
- Recíproca, $X' = 1/X$;
- Exponencial negativa, $X' = \exp(-X)$;

2.2.1.2 Multicolinearidade

É denominado de multicolinearidade o fenômeno de alta correlação entre duas ou mais variáveis explicativas de um modelo de regressão linear múltipla. Este fenômeno é indesejado já que torna o resultado do coeficiente de regressão de uma dada variável explicativa enviesado ao ser influenciado pela adoção ou exclusão de variáveis explicativas do modelo. Portanto, o coeficiente de regressão deixaria de explicar a relação entre a variável dependente e a explicativa (GAZOLA, 2002; NUNES, NETO e FREITAS, 2019).

A multicolinearidade pode ser identificada medindo o coeficiente de correlação entre as variáveis explicativas. O coeficiente de correlação varia entre menos um e um, sendo resultados próximos a: um, indicativos de uma correlação positiva; menos um, indicativos de uma correlação negativa e; zero, indicativos de nenhuma correlação entre as variáveis. Resultados maiores que 0,8 ou 0,9 indicam forte correlação.

Outra maneira para se identificar a multicolinearidade é através da análise da dispersão dos resíduos. Ou seja, identificar uma tendência entre os resíduos da regressão comparado à variável explicativa sinaliza a possibilidade de correlacionamento das variáveis (Gazola, 2002).

Em caso de multicolinearidade existem algumas soluções (GAZOLA,2002; NUNES, NETO e FREITAS, 2019):

- aumentar o tamanho da amostra, capaz de corrigir o problema quando ele é fruto de uma amostra inadequada;
- remover a(s) variável(is) correlacionadas;
- utilizar o modelo de *Ridge Regression* (Regressão a Cumeieira).

2.2.1.3 Homocedasticidade

Variância constante, ou homocedasticidade, refere-se à condição em que a variância dos resíduos não é afetada por valores extremos das variáveis explicativas. Quando ocorre a heterocedasticidade o resíduo não será constante ou aleatório. Este fenômeno pode ser identificado pela análise dos resíduos da regressão comparados aos valores de estimados da variável dependente.

Em casos de heterocedasticidade uma transformação na variável resposta deve ser feita (GAZOLA, 2002). Essas transformações seguem a estrutura mencionada no subcapítulo de linearidade, mas desta vez sobre a variável dependente Y.

2.2.2 Modelo de Preços Hedônicos

Segundo Pinto e Fernandes (2017), os preços hedônicos podem ser definidos como os preços implícitos dos atributos dos imóveis. Para Paixão, Ribeiro e Luporini (2019, p.6): “o modelo de preços hedônicos parte da ideia de que o preço do bem é determinado em função das características”. Na avaliação de imóveis é comum a utilização de regressão linear múltipla utilizando como variáveis explicativas os atributos do imóvel e com o preço sendo a variável dependente (ROSA; OLIVEIRA; PINTO, 2019). Utilizar este método permite a comparação das características agregadas de um conjunto de imóveis e estimar-se, assim, o preço dos seus atributos a partir do coeficiente.

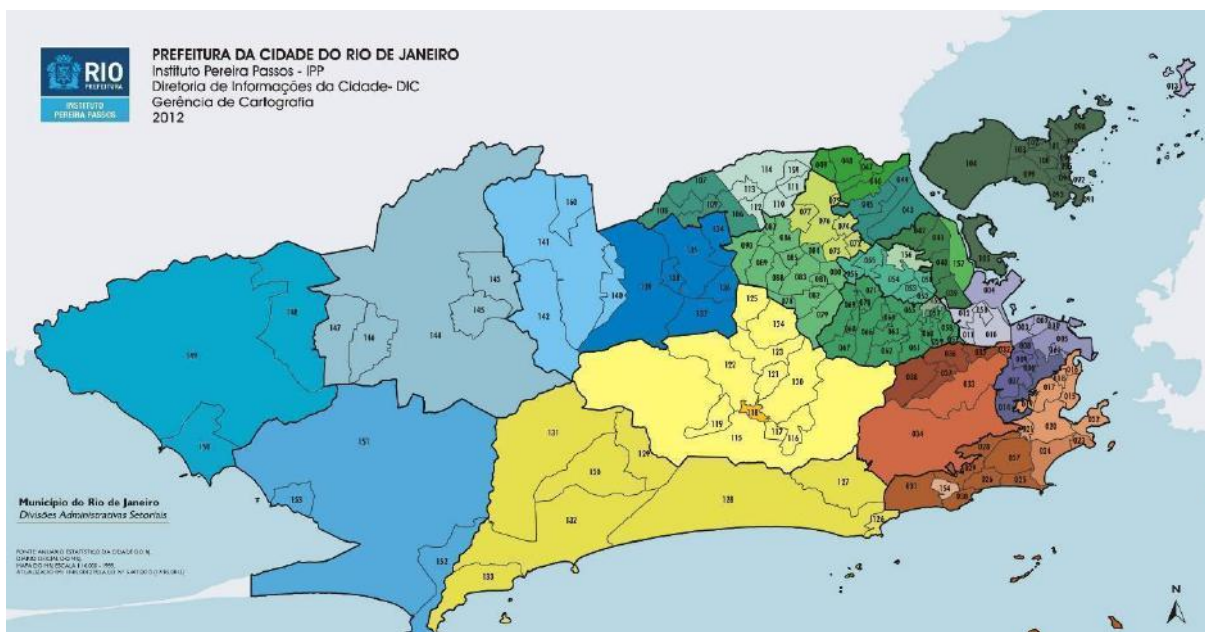
2.3. Visão Geral do Mercado Imobiliário da Cidade do Rio de Janeiro

A cidade do Rio de Janeiro foi fundada no ano de 1565, com o propósito de ocupar o território de domínio português e frustrar a então ocupação francesa da região. A primeira ocupação portuguesa ocorreu na Ilha do Governador e passou para a região do antigo Morro do Castelo e da Praia Vermelha pela importância estratégica desses bairros na entrada da Baía da Guanabara. A cidade tornou-se capital das colônias brasileiras em 1763, título que manteve até 1960 quando a capital do Brasil foi transferida para Brasília.

Atualmente, a cidade é detentora de um território de 1.200 km², dividido em 164

bairros e trinta e três distritos. A Figura 1 é uma representação da demarcação dos bairros e distritos da cidade, a divisão dos distritos é representada pelas cores e os bairros pelas numerações.

Figura 1 - Mapa do município do Rio de Janeiro



Fonte: Prefeitura da cidade do Rio de Janeiro (2012).

De acordo com a pesquisa de Andrade (2016) sobre a condição de moradia no Rio de Janeiro em 2010, foram encontrados que 220.852 residências estão em estado de déficit habitacional, o que corresponde a 10,15% dos domicílios ofertados na cidade.

Segundo o último censo (IBGE, 2010), existem ao menos dois milhões e quatrocentos mil domicílios em toda a cidade do Rio de Janeiro, para uma população de seis milhões e trezentos mil habitantes. A Tabela 1 apresenta o quantitativo dos domicílios no município por tipo e condição de ocupação. Pode-se observar como as casas predominam na cidade, sendo 61% dos domicílios totais e 73% dos domicílios não alugados.

A Tabela 2 apresenta os valores cobrados por metro quadrado em ofertas de imóveis em 2023, com base em dados do site Agente Imóvel (2023). Os dados apresentavam originalmente noventa bairros do município e foram reduzidos para os cinco mais valorizados e cinco menos valorizados. Pode-se observar que os bairros mais caros estão todos localizados na Zona Sul, enquanto os mais baratos estão entre

a Zona Norte e Oeste. Além disso, a amplitude entre a valoração do bairro mais caro comparado com o mais barato chega ao patamar de vinte mil reais.

Tabela 1 - Distribuição dos domicílios por condição de ocupação no município do Rio de Janeiro

Condição de Ocupação	Casa	Apartamento	Cômodo	Total
Próprio	1.015.805	544.236	7.724	1.567.765
Alugado	241.357	225.422	11.884	478.663
Cedido	48.342	33.529	1.507	83.378
Outra Condição	10.069	3.582	988	14.639

Fonte: IBGE (2010).

Tabela 2 - Preço por metro quadrado no Rio de Janeiro por bairros selecionados

Bairro	Preço por m²
Leblon	23.864
Ipanema	22.828
Lagoa	17.294
Jardim Botânico	15.733
Gávea	15.242
Campinho	3.510
Pedra de Guaratiba	3.490
Guaratiba	3.349
Realengo	3.307
Praça Seca	3.165

Fonte: Adaptado dos dados da Agente Imóvel (2023).

3. METODOLOGIA

Este capítulo tem como objetivo explicitar a classificação e método empregados na pesquisa e discorrer sobre a estratégia e técnicas de coleta e análise de dados utilizadas.

3.1. Método, Classificação e Técnicas de pesquisa

Esta é considerada uma pesquisa de abordagem quantitativa, com corte transversal. Segundo Creswell (2007), a pesquisa quantitativa pode ser definida como aquele que emprega o uso de variáveis, com mensuração e estratégias padronizadas para a coleta de dados. Neste estudo, os dados foram obtidos a partir de levantamento.

No quesito de objetivos da pesquisa, esta é considerada uma pesquisa descritiva de acordo com Vergara (1998), pois tem como seu cerne estudar a correlação entre variáveis e explicar um fenômeno a partir delas. Enquanto que sobre a técnica de pesquisa, classifica-se como documental pela utilização de documentos que não passaram por um tratamento e de fonte diversa, neste caso específico, de classificados digitais (VERGARA, 1998).

3.2. Coleta dos Dados

Para a amostra foram coletados dados de anúncios no mercado formal por meio do site Zap Imóveis. As buscas foram limitadas a anúncios de vendas de imóveis residenciais na cidade do Rio de Janeiro. Os anúncios incluem informações sobre seus atributos físicos, locais e o preço ofertado pelo anunciante.

Foi utilizado um algoritmo em Python para fazer uma busca automatizada dos anúncios, buscando assim maximizar o tamanho da amostra. A coleta foi feita de forma automatizada entre os dias 24 e 29 de outubro de 2022. Cada pesquisa feita no site é limitada a um total de 416 (quatrocentos e dezesseis) páginas com 24 (vinte e quatro) anúncios por página. Foram feitas duas pesquisas no site permitindo que fossem coletados 9.984 (nove mil, novecentos e oitenta e quatro) anúncios para a amostra de apartamentos e 9.984 (nove mil, novecentos e oitenta e quatro) para a amostra de casas, antes de qualquer limpeza e exclusão de dados.

Das informações disponíveis no site foram coletados o número de quartos, banheiros e vagas, a área do imóvel, a descrição feita pelo anunciante e o bairro em que o imóvel está localizado. A descrição foi coletada no intuito de servir para criar variáveis relacionadas a amenidades através de uma busca textual. Foi coletado, também, um identificador da base de dados eletrônica para ajudar na identificação de duplicatas.

Tendo em vista que a coleta foi automatizada, a limpeza dos dados tornou-se imprescindível. Foram excluídas as duplicatas e anúncios sem um nome de bairro válido. Também tiveram de ser excluídos casos de anúncios pagos de lançamentos imobiliários através de um filtro textual. Após a limpeza restaram 9.191 (nove mil cento e noventa e um) anúncios de casas e 9.426 (nove mil quatrocentos e vinte e seis) anúncios de apartamentos na amostra.

3.3. Seleção das Variáveis

A escolha das variáveis para esta pesquisa foi baseada nos estudos referenciados. A literatura (LACERDA e FUNCIA, 2005) faz distinção para quatro grupos de características dos imóveis que podem ser determinantes para a formação do seu preço, dois grupos se destacam como os mais utilizados e mais relevantes: as características de localização e as de estrutura física. Representando os atributos físicos dos imóveis foram utilizadas as variáveis de tamanho físico, número de quartos, banheiros, número de vagas e uma *dummy* representando a existência de piscina. Para a localização foram utilizadas *dummies* representando o bairro, região ou região administrativa daquele imóvel. A região administrativa representa a subprefeitura onde aquele imóvel se localiza e funcionou como um intermediário de granularidade entre bairro e região.

3.4. Análise dos dados

Para a análise dos dados foi utilizado o método de regressão múltipla. Segundo Stertz *et al.* (2016) regressão linear múltipla pode ser definida como a técnica estatística que é utilizada para aferir o peso de cada variável explicativa sobre a variável dependente e, posteriormente, a partir desta relação projetar um valor a partir de uma dada coleção de variáveis. Quando esta técnica é utilizada na avaliação

imobiliária, a variável dependente representa o valor do imóvel e as variáveis explicativas representam as características que influenciam na formação do preço (MALAMAN; AMORIM, 2017). A regressão é representada pela Equação 1.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

Sendo o Y a variável dependente ou variável de interesse, x_1 ; x_2 ; x_n as variáveis explicativas ou independentes, β_0 ; β_1 ; β_2 ; β_n os regressores ou coeficiente de regressão, que representam o peso de cada variável explicativa sobre o resultado da formação do valor da variável de interesse Y. O coeficiente é estimado através da minimização da soma dos quadrados dos erros pelo método dos Mínimos Quadrados Ordinários (MQO). Por último vem ε que representa o resíduo da regressão ou erro estatístico.

4. RESULTADOS

Neste capítulo são descritos os resultados encontrados pela pesquisa e uma descrição dos dados coletados.

4.1. Estatística Descritiva

Os dados da Zap Imóveis sobre a cidade do Rio de Janeiro mostraram-se diversos na categoria dos imóveis, com representação de imóveis de baixa até alta renda e de 139 diferentes bairros do Rio de Janeiro entre 18.617 observações. A Tabela 3 apresenta os dez bairros mais representados pela amostra e o número de observações de cada um. A escolha por esse tipo de coleta teve como objetivo maximizar o tamanho da amostra e testar a capacidade de estudos com uma grande amostra de dados de classificados online.

Tabela 3 - Dez Bairros mais representados pela amostragem

Bairro	Região	Contagem total	Número de Apartamentos
Barra da Tijuca	Zona Oeste	1687	945
Tijuca	Zona Norte	1374	881
Recreio dos Bandeirantes	Zona Oeste	1332	731
Copacabana	Zona Sul	1329	1257
Campo Grande	Zona Oeste	972	49
Botafogo	Zona Sul	928	701
Ipanema	Zona Sul	526	496
Taquara	Zona Oeste	501	165
Laranjeiras	Zona Sul	393	220
Leblon	Zona Sul	393	372

Fonte: elaborado pelo autor (2023).

Nota-se que os mais representados são principalmente da Zona Sul e Zona Oeste, sendo a Tijuca o único bairro da Zona Norte entre eles e nenhum bairro do

centro da cidade. A Zona Central da cidade, apesar de reformas recentes do poder municipal, não apresenta uma oferta considerável de imóveis à venda, com apenas 268 observações em toda a região.

As principais limitações dessa coleta foram: a falta e, por vezes, má organização das informações em alguns anúncios; e o fato dos preços apenas representarem a oferta inicial e, por isso, possivelmente não corresponderem ao preço praticado no ato de negociação.

A Tabela 4 apresenta a estatística descritiva apenas dos apartamentos. A tabela foi restringida às características intrínsecas e locacionais relacionadas à região contidas no modelo, desconsiderando as *dummies* de bairro e região administrativa. A amostra deste recorte contém 9.426 (nove mil quatrocentos e vinte e seis) anúncios de apartamentos.

Tabela 4 - Estatística Descritiva do recorte final dos dados

Variáveis	média	desvio padrão	mínimo	mediana	máximo
Preço	1.257.267	1.566.119	75.000	802.000	28.307.597
Area	108	70,99	10	87	1.200
Quarto	2,55	0,84	1	3	8
Banheiro	2,26	1,21	1	2	15
Vaga	0,58	1,22	0	0	45
Zona Sul	0,4372	0,50	0	0	1
Zona Norte	0,28	0,45	0	0	1
Zona Central	0,02	0,14	0	0	1
Zona Oeste	0,27	0,44	0	0	1
Piscina	0,17	0,37	0	0	1

Fonte: elaborado pelo autor (2023).

Esta tabela representa os dados como se encontram no final dos testes, com um filtro excluindo as casas. O preço médio é de 1,257 milhões de reais, o desvio padrão elevado indica que os dados estão distantes da média. Isso é reforçado pela

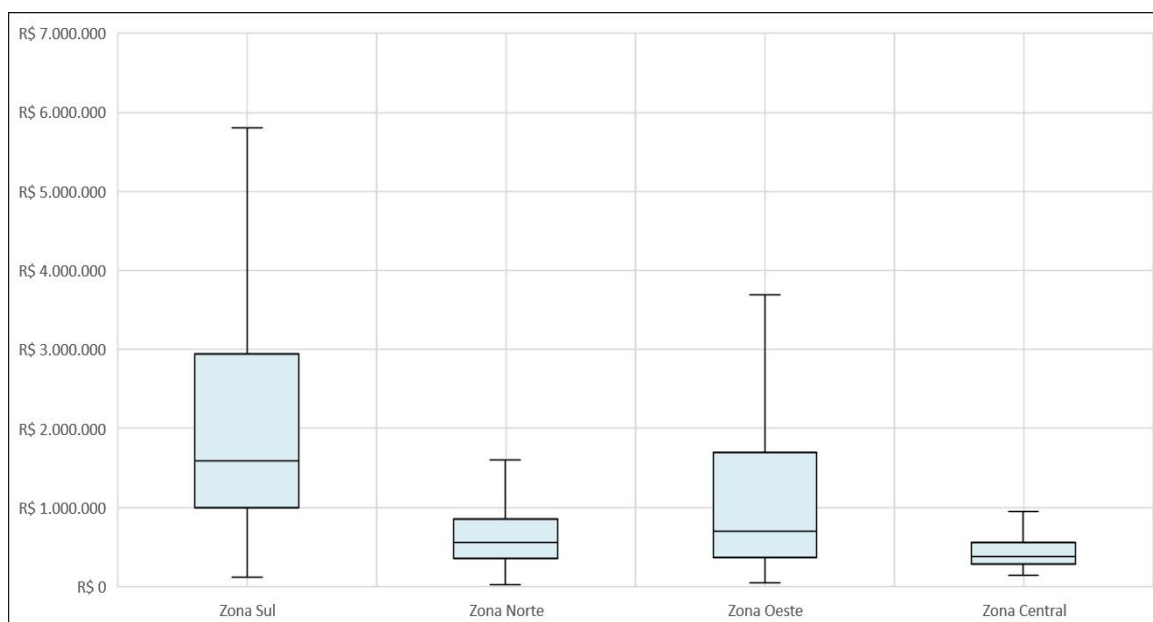
mediana que diz que metade das observações se posicionam abaixo da faixa de 802 mil.

Sobre as variáveis que representam os atributos físicos dos imóveis, nota-se que os imóveis têm, em média, 10 m² de área, entre dois e três quartos, dois banheiros, e entre nenhuma e uma vaga. Destas variáveis a área e número de vagas se destacam pela grande variedade entre o mínimo e o máximo, com um elevado desvio padrão.

Nos Gráficos 1 e 2 pode-se observar a distribuição dos preços da amostra representada em *boxplots* sem os *outliers*, que foram excluídos para preservar a escala. A linha média do *boxplot* é a mediana, os limites inferiores e superiores da caixa são o primeiro e o terceiro quartil e os traços representam os limites inferiores e superiores da distribuição.

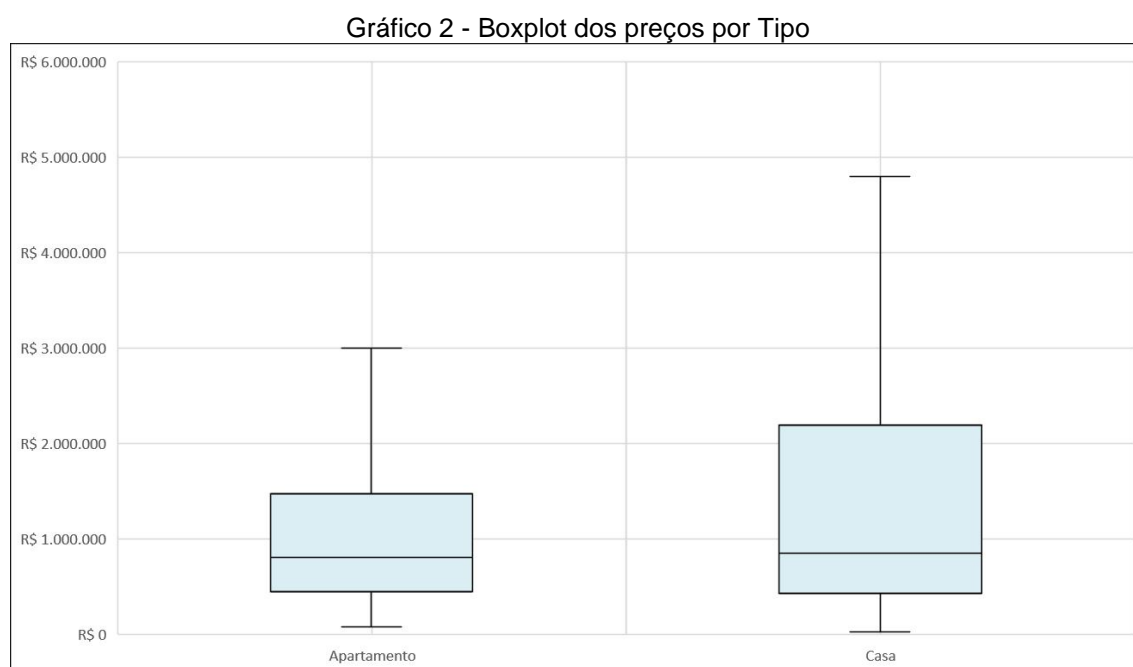
No Gráfico 1, o mínimo em todas as categorias é similar, demonstrando que a oferta de imóveis de baixo valor é consistente por toda a cidade. As regiões com menor amplitude na variação dos preços são as regiões Norte e Central. Em contraposição, a região Sul e Oeste apresentam maiores variações entre os limites inferiores e superiores do *boxplot*. Isto é explicado pelo fato da região Sul abrigar os bairros mais nobres e famosos da cidade como Leblon, Copacabana e Ipanema, enquanto a região Oeste contém a Barra da Tijuca, um bairro que teve forte valorização nas últimas décadas.

Gráfico 1 - Boxplot dos preços por Região



Fonte: elaborado pelo autor (2023).

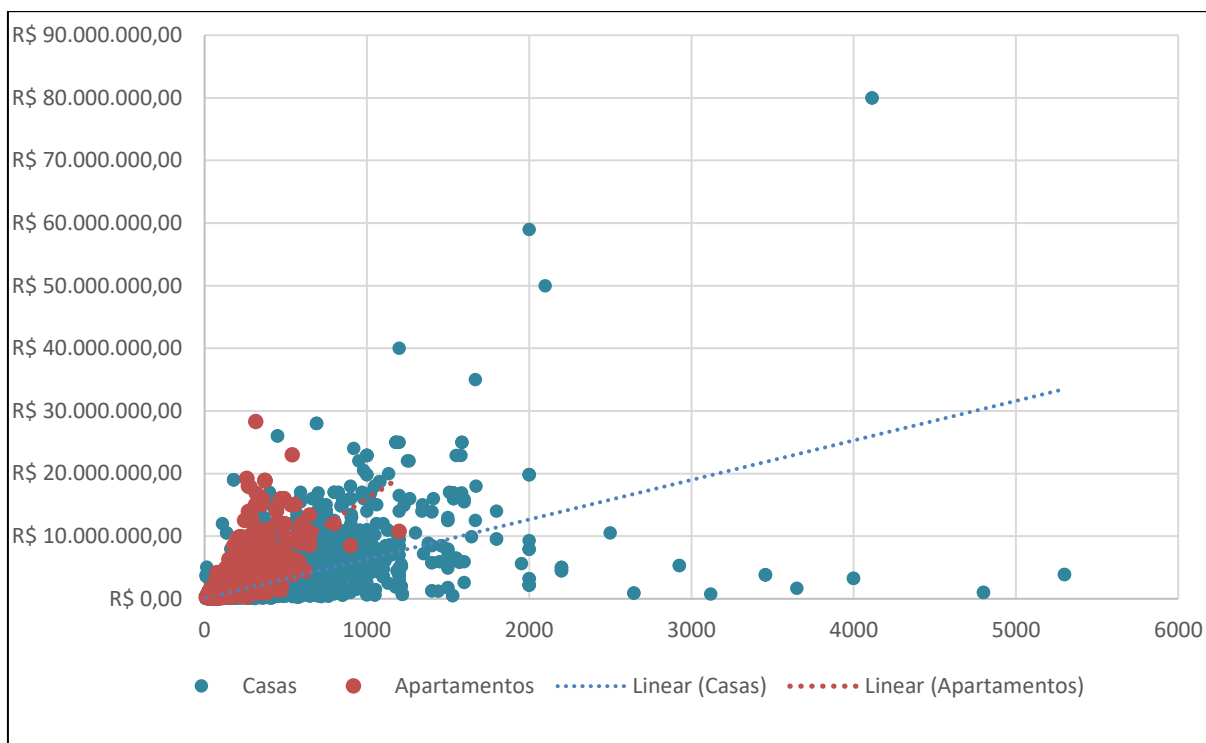
No Gráfico 2 é evidente a maior amplitude no preço das casas do que de apartamentos. Isto não significa que casas são mais valorizadas em geral, evidenciado pela mediana em ambos *boxplots* serem próximas de 800 mil e seus primeiros quartis também estarem próximos com 445 mil no caso dos apartamentos e 430 mil para as casas. Na amostra, essa maior amplitude no preço das casas é causada pelo fato da oferta de casas de luxo ter um preço mais elevado que o preço dos apartamentos de luxo.



Fonte: elaborado pelo autor (2023).

A relação preço e área dos imóveis é evidente no Gráfico 3. Neste gráfico as observações foram classificadas de acordo com o tipo de imóvel e é possível observar uma linha de tendência para cada um. Também foi necessário excluir do gráfico três observações com área maior que dez mil metros quadrados para permitir uma melhor visualização. A tendência mostra uma diferença considerável quanto ao tipo (casa ou apartamento), o que evidencia que esta variável tem uma influência maior no preço quando o imóvel é um apartamento e uma influência menor no caso das casas.

Gráfico 3 – Dispersão dos preços por Metro quadrado



Fonte: elaborado pelo autor (2023).

4.2. Regressão Linear Múltipla

Este subcapítulo apresenta a aplicação prática do modelo para determinação da equação de regressão, começando com a determinação das variáveis utilizadas e os três modelos escolhidos para o trabalho. Após, são testados os modelos para saber qual tem maior capacidade explicativa. O melhor modelo, então, é testado pelos pressupostos da regressão linear, sendo necessário fazer transformações nas variáveis para adequação do modelo.

4.2.1 Equação do Modelo

O método escolhido para fazer a regressão foi o método linear múltiplo devido a sua capacidade de correlacionar diferentes variáveis a uma variável dependente. Durante a pesquisa, foram usados diferentes recortes dos dados, aqui chamados modelos até chegar no modelo final, que apresentou melhor resultado de R-Quadrado Ajustado. As Equações 2, 3 e 4 representam os três recortes iniciais.

$$\gamma_{preço} = \beta_0 + \beta_1 Q + \beta_2 A + \beta_3 BN + \beta_4 VG + \beta_5 PS + \beta_6 CA + \alpha_1 ZN + \alpha_2 ZS + \alpha_3 ZC + \alpha_4 ZO + \varepsilon \quad (2)$$

$$\gamma_{preço} = \beta_0 + \beta_1 Q + \beta_2 A + \beta_3 BN + \beta_4 VG + \beta_5 PS + \beta_6 CA + \alpha_1 RA_1 \cdots + \alpha_n RA_n + \varepsilon \quad (3)$$

$$\gamma_{preço} = \beta_0 + \beta_1 Q + \beta_2 A + \beta_3 BN + \beta_4 VG + \beta_5 PS + \beta_6 CA + \alpha_1 BR_1 \cdots + \alpha_k BR_k + \varepsilon \quad (4)$$

Das variáveis presentes nos três modelos:

- Y representa o valor ofertado pelo imóvel
- Q é o número de dormitórios
- A é a área apresentada no anúncio
- BN é o número de banheiros
- VG é a quantidade de vagas
- PS é uma variável dummy que assinala quando o imóvel tem uma piscina
- CA é outra dummy que assinala se o imóvel é uma casa

As variáveis restantes diferenciam os recortes uns dos outros. No primeiro modelo foi feito um recorte por região, portanto ZN , ZS , ZC e ZO são variáveis dummies que representam em que região o imóvel está localizado: Zona Norte, Zona Sul, Zona Central e Zona Oeste, respectivamente. No segundo modelo RA_1 até RA_n representam a região administrativa e, por último, BR_1 até BR_k representam os bairros.

4.2.2 Resultados

As estimações de regressão foram feitas neste trabalho utilizando a linguagem de programação *Python* e as bibliotecas: *statsmodelapi*, para os cálculos dos coeficientes da regressão; *scikitlearn* e *numpy* para transformações das variáveis. O procedimento adotado foi estimar a equação com todas as variáveis dos diferentes modelos, para depois fazer a seleção do modelo com melhor resultado e, então, fazer uma seleção das variáveis significantes do modelo. No final, ainda foram testados os pressupostos de regressão linear múltipla.

A Tabela 5 apresenta a primeira estimação dos parâmetros, nela a amostra de casa apresentou um resultado muito baixo para o R-Quadrado Ajustado, com um

resultado de 0,356. Em contraposição, a primeira estimativa para a amostra de apartamentos demonstrou um R-Quadrado Ajustado de 0,621.

A causa dessa diferença pode ter sido uma menor qualidade dos dados para as casas, como por exemplo uma falta de distinção entre área do imóvel e área construída no preenchimento do anúncio ou exagero no número de vagas para automóveis. Com uma análise de algumas das observações da amostra foi possível observar um padrão onde a área anunciada na oferta era menor que a descrita no texto. Outra causa provável é esse preço ser explicado por variáveis que não estão presentes como uma razão entre área construída e área total do imóvel, que em algumas observações foi confundida pelo anunciante. Por essas razões, a amostra de casas se mostrou insuficiente em qualidade para continuar no modelo.

Tabela 5 – Resultados regressão do recorte de região por tipo de imóvel

Amostra	R2	R2-Ajustado
Casas	0,357	0,356
Apartamentos	0,622	0,621
Casas e Apartamentos	0,373	0,372

Fonte: elaborado pelo autor (2023).

Com os dados restringidos a apenas os apartamentos, foi gerado a Tabela 6. Essa tabela demonstra a variância obtida pelo modelo de regressão para a amostra de 9.426 apartamentos. O recorte por região geográfica apresentou o menor valor para a variância em 62%, o que significa que a equação do modelo foi capaz de explicar 62% do valor. Como era esperado, os recortes por região administrativa e bairros tiveram maior capacidade explicativa por terem mais informações sobre a localização dos imóveis.

A Equação 3 teve a maior capacidade explicativa e, então, foi escolhida para o modelo. O valor máximo para R-Quadrado Ajustado em 0,69 não é muito alto em comparação a outros trabalhos, porém é satisfatório tendo em consideração que este trabalho teve como objeto de estudo uma cidade heterogênea como o Rio de Janeiro.

Tabela 6 - Resultados por recorte geográfico

Recorte	R2	R2-Ajustado
Região	0,622	0,621
Região Administrativa	0,676	0,675
Bairros	0,695	0,691

Fonte: elaborado pelo autor (2023).

O Apêndice. A apresenta o coeficiente estimado de cada variável e o p-valor. O p-valor menor que 0,05 indica que podemos rejeitar a hipótese nula de que remover a variável não trará consequência significativas ao modelo. Os dados estão organizados de acordo com os menores p-valor, para manter as variáveis mais relevantes no topo. É possível observar pelos coeficientes que os bairros da Zona Sul, como: Copacabana, Laranjeiras, Leblon, Gávea, tendem a ter um preço elevado devido a sua localização. Em contraposição, bairros de outra região foram desvalorizados como Praça Anchieta, Recreio dos Bandeirantes e Tijuca, com exceção da Barra da Tijuca. Todas as variáveis essenciais para a modelagem como área, número de banheiros, quartos e vagas se mostraram relevantes ao analisar o p-valor.

4.2.3 Verificação dos Pressupostos

Com os primeiros testes concluídos e a equação do modelo determinada, foram feitos testes para verificar e adequar o modelo aos pressupostos da regressão.

4.2.3.1 Homocedasticidade

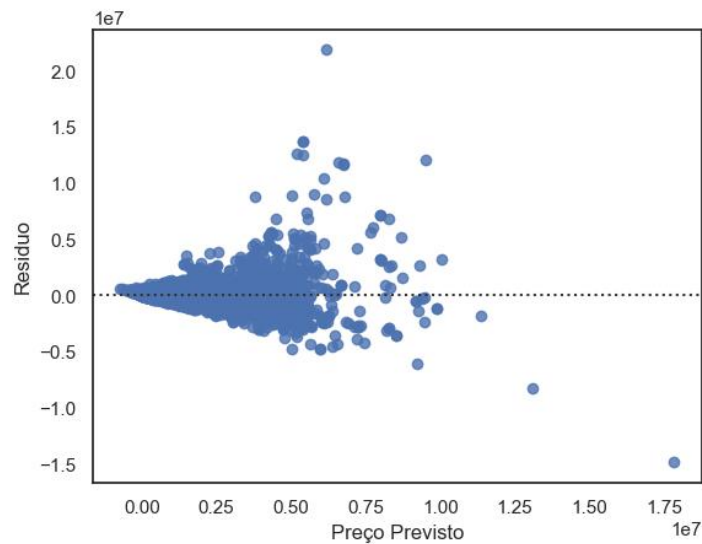
A homocedasticidade, ou variância constante dos erros, é constatada pela análise gráfica dos resíduos e os valores estimados para a variável dependente. No Gráfico 6 é feita esta comparação e pode ser constatado que o modelo não atende este pressuposto, dada a diferença de dispersão dos resíduos.

Constatado este fato, torna-se necessário fazer uma transformação da variável dependente e averiguar novamente este pressuposto. Optou-se por fazer a

transformação logarítmica da variável (Equação 5), pois ela é a preferida para ajustes na variável dependente no estudo do mercado imobiliário (GAZOLA, 2002).

$$Y = \ln(Y_{preço}) \quad (5)$$

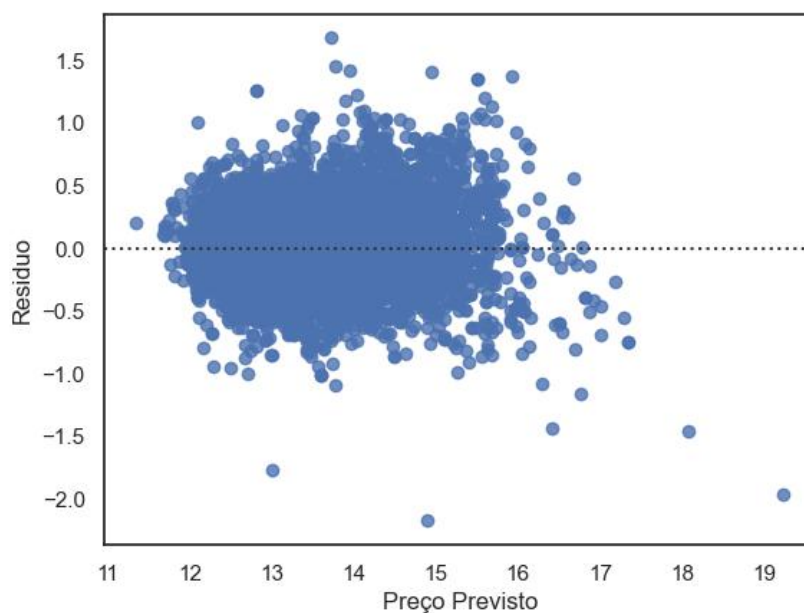
Gráfico 4 - Resíduos por preço estimado



Fonte: elaborado pelo autor (2023).

O Gráfico 5 demonstra que a transformação obteve êxito em tornar os erros aleatórios. Então é possível afirmar que o modelo atende ao pressuposto de homocedasticidade.

Gráfico 5 - Resíduos por preço estimado após transformação



Fonte: elaborado pelo autor (2023).

4.2.3.2 Colinearidade

A Figura 2 apresenta o coeficiente de correlação entre as variáveis explicativas quantitativas. É aparente uma forte correlação entre a variável área e quartos em 0,7. De acordo com Gonzalez (2002) e a Norma (ABNT, 2011), tratamentos especiais devem ser utilizados para casos de correlação acima de 0,8, o que não ocorreu.

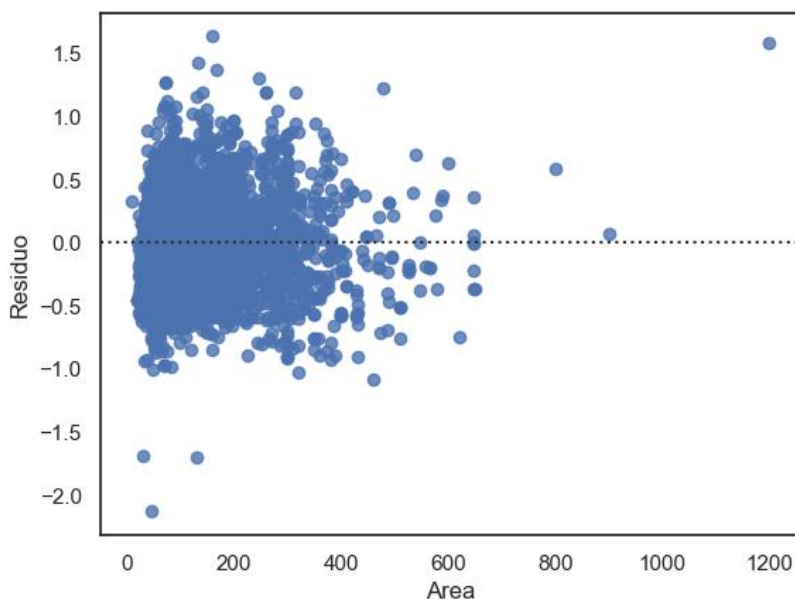
Figura 2 - Coeficiente de correlação das variáveis quantitativas do modelo



Fonte: elaborado pelo autor (2023).

No Gráfico 6 é feita a análise dos resíduos padronizados em comparação à variável área. Como não é possível encontrar um padrão na análise gráfica, a variável foi mantida no modelo na condição de inexistência de colinearidade.

Gráfico 6 - Resíduo por Área



Fonte: elaborado pelo autor (2023).

4.3. Seleção de Variáveis

A partir dos primeiros resultados foi necessária uma nova seleção de variáveis a fim de restringir o número de variáveis do modelo de 128 para apenas 15 variáveis, mantendo apenas as mais significantes. Para executar esta seleção foram utilizados o método de seleção progressiva de variáveis e o de seleção regressiva com base na maximização do R-quadrado.

De acordo com Araújo *et al.* (2010), a seleção progressiva, ou *stepwise*, é um método de seleção de variáveis que utiliza um algoritmo para fazer a regressão em etapas. Começando com apenas uma variável independente e inserindo a cada etapa uma nova variável, até que a inclusão de novas variáveis não produza uma melhora no desempenho. No caso deste trabalho o desempenho foi medido pelo R-quadrado Ajustado.

A Tabela 7 apresenta as variáveis selecionadas pela seleção progressiva por ordem, com as mais significantes primeiro e as menos significativas por último. Entre as variáveis selecionadas, a com maior poder explicativo foi a área, que corresponde a mais da metade da variância. Também se destaca como foi excluído o número de vagas do modelo. Da maioria das variáveis selecionadas muitas são de bairros nobres da região sul do município como Ipanema, Leblon, Flamengo, Botafogo, Laranjeiras e outros. As únicas exceções entre as variáveis locais selecionadas foram a Barra

da Tijuca e Recreio dos Bandeirantes, ambos bairros da Zona Oeste com grande oferta de imóveis.

Tabela 7 - Variáveis selecionadas pela seleção progressiva

Número de Variáveis	Variável Adicionada	R-Quadrado
1	área	0,521
2	Ipanema	0,568
3	Leblon	0,620
4	banheiros	0,653
5	Copacabana	0,677
6	Botafogo	0,706
7	Lagoa	0,729
8	Barra da Tijuca	0,755
9	Flamengo	0,768
10	Jardim Botânico	0,781
11	Laranjeiras	0,793
12	Gávea	0,802
13	quartos	0,812
14	Humaitá	0,821
15	Recreio dos Bandeirantes	0,830

Fonte: elaborado pelo autor (2023).

Em contrapartida, a seleção regressiva parte dos mesmos princípios de resolver a equação de regressão em etapas e medindo seu desempenho. Porém, ele parte da equação com todos os regressores e, a cada iteração, retira uma variável e compara se impactou negativamente o resultado. Caso não ocorra uma mudança negativa no desempenho, o passo é mantido e o processo segue até se exaurir as variáveis a retirar.

Na Tabela 8 estão representadas todas as quinze variáveis selecionadas pela seleção regressiva. Por esse método alcançou-se um R-Quadrado de 0,833.

Comparado à seleção progressiva muitas das variáveis se repetem, novamente foi excluído a variável para número de vagas. Porém, dessa vez foi selecionado a variável que representa o bairro da Tijuca.

Tabela 8 - Variáveis selecionadas pela seleção regressiva

Variáveis Selecionadas
área
quartos
Flamengo
Copacabana
Tijuca
Botafogo
Barra da Tijuca
Laranjeiras
Lagoa
Recreio dos Bandeirantes
Leblon
Jardim Botânico
Ipanema
Humaitá
Gávea

Fonte: elaborado pelo autor (2023).

4.4 Discussão sobre os resultados

Para concluir este capítulo foram feitos novos testes incluindo as variáveis selecionadas por ambos os métodos de seleção, como representado pela Equação 6. A Equação 7 representa o cálculo do preço do imóvel a partir da Equação 6.

Todas as variáveis rejeitam a hipótese nula de acordo com o p-valor, assim como pode-se observar na Tabela 9. Também é notável a influência das variáveis

locacionais sobre o preço com seus coeficientes elevados. Esses coeficientes apresentam os cinco bairros mais valorizados na mesma ordem que a Tabela 2.

$$\ln(Y_{preço}) = \beta_1 Q + \beta_2 A + \beta_3 BN + \beta_4 Flamengo + \beta_5 Copacabana + \beta_6 Tijuca + \beta_7 Botafogo + \beta_8 Barra da Tijuca + \beta_9 Laranjeiras + \beta_{10} Lagoa + \beta_{11} Recreio dos Bandeirantes + \beta_{12} Leblon + \beta_{13} Jardim Botânico + \beta_{14} Ipanema + \beta_{15} Humaitá + \beta_{16} Gávea + \varepsilon \quad (6)$$

$$Y_{preço} = \exp[\beta_1 Q + \beta_2 A + \beta_3 BN + \beta_4 Flamengo + \beta_5 Copacabana + \beta_6 Tijuca + \beta_7 Botafogo + \beta_8 Barra da Tijuca + \beta_9 Laranjeiras + \beta_{10} Lagoa + \beta_{11} Recreio dos Bandeirantes + \beta_{12} Leblon + \beta_{13} Jardim Botânico + \beta_{14} Ipanema + \beta_{15} Humaitá + \beta_{16} Gávea + \varepsilon] \quad (7)$$

Tabela 9 - Coeficiente Estimados

Variável	Coeficiente Estimado	p-valor
Leblon	13,61	0,0E+00
Ipanema	13,53	0,0E+00
Lagoa	13,32	0,0E+00
Jardim Botânico	13,25	1,5E-194
Gávea	13,23	1,8E-131
Botafogo	13,03	0,0E+00
Humaitá	12,99	2,6E-123
Copacabana	12,95	0,0E+00
Flamengo	12,93	1,0E-239
Barra da Tijuca	12,86	0,0E+00
Recreio dos Bandeirantes	12,46	5,6E-122
Tijuca	12,42	1,1E-92
quartos	0,1368	2,6E-94
banheiros	0,0559	2,2E-92
área	0,0042	0,0E+00

Fonte: elaborado pelo autor (2023).

O modelo final apresentou um coeficiente de determinação R^2 em 83%, mesmo patamar atingido pelos estudos de Medeiros e Carvalho (2017) e de Fernandes e Pinto (2019) que utilizaram do mesmo método de regressão e menor que o trabalho de Gazola (2002) que fez uso da *Ridge Regression* e alcançou previsibilidade de 96%.

No resultado final, o trabalho aponta que características intrínsecas como tamanho do imóvel e a existência de uma piscina tem uma influência menor no preço de venda do bem imobiliário quando comparado ao efeito de vizinhança. Como foi encontrado nos trabalhos de Fernandes e Pinto (2019) e Rosa, Oliveira e Pinto (2019), a localização do imóvel está relacionada ao acesso do morador a bens de infraestrutura, lazer, cultura e consumo. Esta valorização também pode estar relacionada a preferência de segmentos mais abastados por imóveis nas regiões representadas como mais valorizadas (RIBEIRO e LUPORINI, 2019). Como pode ser observado o modelo assinalou como mais valorizados os bairros da Zona Sul, Leblon até Flamengo, seguidos da Barra da Tijuca e num patamar abaixo, os bairros do Recreio dos Bandeirantes e Tijuca.

Entre as características intrínsecas ao imóvel as mais relevantes para a formação do preço foram número de quartos, banheiros e área do imóvel. A significância dessas variáveis é previsível e indica uma relação positiva entre o tamanho da unidade imobiliária e seu preço de venda.

5. CONCLUSÃO

Este trabalho teve como objetivo mapear as diferentes variáveis que afetam o preço dos imóveis no mercado imobiliário carioca como um todo. Como resultado apresentou que a localização é o maior diferencial na formação dos preços, enquanto suas características físicas têm uma influência menor, mas ainda importante. Este resultado sugere que o ajuste de modelos específicos para as diferentes localizações pode ser necessário para melhorar a representação dos preços no mercado imobiliário do Rio de Janeiro. Em estudos para mapear a preferência dos consumidores fazer um recorte por segmento e localização pode ser essencial.

O modelo apresentou bons resultados para os testes teóricos dos pressupostos de regressão linear, mas apresentou forte multicolinearidade como era de se esperar dos dados deste mercado (GAZOLA, 2002). É sugerido que próximos trabalhos utilizem técnicas para sanar este problema, como a utilização do método de *Ridge Regression* (Regressão Cumeeira) (GAZOLA, 2002; NUNES, NETO e FREITAS, 2019).

O trabalho exposto pode contribuir para a compreensão do mercado imobiliário carioca e diminuir a assimetria de informações para investidores do ramo. Os resultados evidenciam as subregiões valorizadas da cidade e características intrínsecas que influenciam no preço de venda do imóvel. Neste sentido, este trabalho pode auxiliar investidores a avaliarem preços de venda. Entretanto, o modelo tem de ser ajustado com periodicidade para manter-se atualizado à dinâmicas do mercado.

Quanto a técnicas utilizadas, o trabalho contribuiu quanto a utilização de meios automatizados para a criação do banco de dados. A utilização de extração de dados da internet auxiliou na criação de um grande banco de dados, mas trouxe alguns reveses. Os anúncios para casas tinham um padrão de má apresentação dos dados pelos ofertantes e o grande número de informações tornou impossível um saneamento caso a caso desses dados. Um ponto que poderia ter sido expandido no quesito dos dados, seria uma utilização de algoritmos para extrair novas variáveis a partir dos textos da descrição, sendo esta uma sugestão para trabalhos futuros.

REFERÊNCIAS

- ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 14653-2: Avaliação de bens – Parte 2: Imóveis urbanos. Rio de Janeiro, 2011.
- ALVES, D. *et al.* Modelagem dos preços de imóveis residenciais paulistanos. **Revista Brasileira de Finanças**, v. 9, n. 2, p. 167-187, 2011.
- ANDRADE, E. **Necessidades Habitacionais no Rio de Janeiro**. PCRJ/SMU/CGPU/Coordenadoria de Macroplanejamento, 2016.
- ARAÚJO, E. *et al.* Proposta de uma metodologia para a avaliação do preço de venda de imóveis residenciais em Bonito/MS baseado em modelos de regressão linear múltipla. **P&D em Engenharia de Produção**, Itajubá, v. 10, n. 2, p.195-207, 2012.
- ARRAES, R. A.; FILHO, E. S. **Externalidades e formação de preços no mercado imobiliário urbano brasileiro**: um estudo de caso. *Economia Aplicada*, v. 12, n. 2, p. 289-319, abr./ jun. , 2008.
- BRANDO, L.; BARBEDO, C. H. Há Fatores Não Econômicos na Formação do Preço de Imóveis?. **Revista de Administração Contemporânea**. v. 20, n. 1, p. 106-130, 2016. Disponível em: <<https://doi.org/10.1590/1982-7849rac2016140095>>. Acesso em: 01 ago. 2022.
- CRESWELL, J. W. **Procedimentos qualitativos. Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 2ed. Porto Alegre: Artmed, 2007
- DIEWERT, W. E. **The Paris OECD-IMF workshop on real estate price indexes: Conclusions and Future Directions**. In: DIEWERT, E. Erwin et al. (Ed.). *Price and Productivity Measurement: volume 1 – housing*. Trafford Press, 2009.
- FIPE . **FipeZAP**: Índice Fipezap de Preços De Imóveis Anunciados. 2022. Disponível em:<<https://www.fipe.org.br/pt-br/publicacoes/relatorios/>> . Acesso em: 11 mai. 2022.
- FERNANDES, A. M. R.; MOREIRA, D. S.; SILVA, R. S. Utilização de um Rede Neural Artificial e Análise Multicritério para determinação do valor de aluguel de apartamentos. *In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA*, 7., 2010, Resende, RJ. **Anais [...]**. Disponível em: <https://www.aedb.br/seget/arquivos/artigos10/253_seget%20todim.pdf>. Acesso: 16 jun. 2022
- GAZOLA, S. **Construção de um modelo de regressão de avaliação de imóveis**. 2002. 104 f. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis.
- IBGE. **Censo Demográfico**, 2010. Disponível em <<https://censo2010.ibge.gov.br/sinopse/index.php?uf=33&dados=29>>. Acesso em: 14 de mai 2022.

LACERDA, F. C. ; FUNCIA, C. O impacto da violência criminal urbana no preço dos imóveis residenciais na região da Tijuca, cidade do Rio de Janeiro: um estudo exploratório. **Cadernos EBAPE.BR** , v. 3, n. 3 , p . 01-15, 2005. Disponível em: <<https://doi.org/10.1590/S1679-39512005000300007>>. Acesso em: 1 ago. 2022.

MALAMAN, C.; AMORIM, A.. Método para determinação de valores na avaliação imobiliária : comparação entre o Modelo de Regressão Linear e Lógica Fuzzy. **Boletim de Ciências Geodésicas** , v. 23, n. 1, p. 87-100, 2017. Disponível em: <<http://dx.doi.org/10.1590/s1982-21702017000100006>>. Acesso em: 9 jul. 2022

MEDEIROS, R. ; CARVALHO, S. T. Modelagem econométrica do preço de aluguéis de apartamentos na cidade de Petrópolis-RJ utilizando regressão linear múltipla . **Revista de Economia da UEG**, v. 13 ,n. 1, p . 158-174, 2017.

NUNES, D. B.; BARROS, J. P.; FREITAS, S. M. Modelo de regressão linear múltipla para avaliação do valor de mercado de apartamentos residenciais em Fortaleza, CE. **Ambiente Construído**, v. 19, n. 1, p. 89-104, 2019. Disponível em: <<https://doi.org/10.1590/s1678-86212019000100295>>. Acesso em: 1 ago. 2022.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. Instituto Pereira Passos. **Município do Rio de Janeiro**: Regiões de Planejamento, Regiões Administrativas e Bairros. Rio de Janeiro, 2012. 1 mapa, color. Escala 1:100.000. Disponível em: <https://www.data.rio/documents/PCRJ::regi%C3%B5es-de-planejamento-rp-regi%C3%B5es-administrativas-ra-e-bairros-do-munic%C3%ADpio-do-rio-de-janeiro/explore>. Acesso em 22 mai. 2023.

PINTO, V. H. L.; FERNANDES, R. A. S. Análise de preços hedônicos no mercado imobiliário residencial de Conselheiro Lafaiete, MG. **Interações**, Campo Grande, v. 20, n. 2, p. 627-643, abr./jun. 2019. Disponível em: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1518-70122019000200627>. Acesso em: 01 jul. 2022.

RIBEIRO, A.; LUPORINI, V. A valorização imobiliária em Belo Horizonte, 1995-2012: uma análise hedônica-quantílica. **Nova Economia**. v. 29, n. 3 , p . 851-880, 2019. Disponível em: <<https://doi.org/10.1590/0103-6351/3637>> Acesso em: 9 jul. 2022.

ROSA, V. S.; OLIVEIRA, P. B. de; PINTO, R. L. M. Modelos de precificação para locação e venda de imóveis residenciais na cidade de João Monlevade-MG via regressão linear multivariada. **Gestão da Produção, Operações e Sistemas**, Bauru, v. 14, n. 3, p. 151-167, 2019. Disponível em: <<https://revista.feb.unesp.br/index.php/gepros/article/view/2614>>. Acesso em: 8 jul. 2022.

STERTZ, E. *et al.* Uma análise sobre o comportamento dos preços dos imóveis na cidade de Porto Alegre/RS **GESTÃO.ORG** , v. 14, n. 1, 2016 .

VERGARA, S. **Projetos e Relatórios de Pesquisa em Administração** ed. 2 São Paulo: Editora Atlas, 1998.

APÊNDICE A - Tabela com todos os coeficientes de regressão da Equação 4

Variável	beta estimado	p-valor
Área	14.537	0,00
Leblon	1.969.393	0,00
Ipanema	1.679.554	0,00
Lagoa	1.151.126	0,00
Vagas	105.753	0,00
Quartos	-160.005	0,00
Constante	-406.224	0,00
Botafogo	411.962	0,00
Jardim Botânico	714.010	0,00
Copacabana	275.603	0,00
Banheiros	66.368	0,00
Gávea	687.716	0,00
Recreio dos Bandeirantes	-237.615	0,00
Jardim Guanabara	-542.285	0,00
Humaitá	390.264	0,00
Urca	1.087.440	0,00
Piscina	101.288	0,00
Barra da Tijuca	172.585	0,00
Parque Anchieta	-2.793.149	0,00
Laranjeiras	209.502	0,00
Flamengo	187.614	0,00
Tijuca	-124.296	0,00
Vargem Grande	-625.767	0,01
Centro	205.609	0,02
Vila Isabel	-156.339	0,03
São Conrado	288.789	0,04

Variável	beta estimado	p-valor
Grajau	-175.951	0,05
Vargem Pequena	-366.289	0,06
Santo Cristo	256.785	0,08
Catumbi	-1.064.704	0,08
Leme	227.359	0,09
Catete	234.805	0,10
Vila da Penha	-224.344	0,11
Vista Alegre	-485.114	0,11
Bancários	-441.625	0,13
Praça Seca	-170.886	0,17
Rocha	-404.434	0,19
Méier	-89.625	0,19
Cocota	-291.996	0,23
Taua	-250.877	0,27
Olaria	-212.178	0,30
Inhoaíba	359.570	0,31
Engenho Novo	-109.756	0,31
Estacio	-349.931	0,32
Gloria	123.562	0,35
Cascadura	-224.879	0,35
Andaraí	-100.605	0,35
Jardim America	-319.198	0,37
Pechincha	-64.763	0,41
São Cristovão	122.769	0,41
Campinho	-190.181	0,45
Penha	-112.825	0,50
Higienópolis	-208.259	0,53

Variável	beta estimado	p-valor
Riachuelo	-142.975	0,53
Bonsucesso	129.433	0,54
Agua Santa	-361.201	0,55
São Francisco Xavier	-114.525	0,59
Iraja	-63.173	0,60
Alto da Boa Vista	-180.561	0,61
Penha Circular	-128.077	0,61
Del Castilho	77.641	0,63
Jacarepagua	30.508	0,64
Ribeira	109.855	0,64
Ramos	-108.993	0,64
Itanhangá	-151.972	0,64
Campo Grande	-55.899	0,66
Madureira	-87.810	0,69
Camorim	61.083	0,70
Piedade	-53.588	0,70
Bento Ribeiro	-121.070	0,71
Vidigal	-221.811	0,72
Rio Comprido	-48.551	0,72
Campo dos Afonsos	139.250	0,72
Realengo	-95.052	0,73
Gamboa	146.822	0,73
Praça da Bandeira	53.484	0,74
Anchieta	165.723	0,74
Colégio	200.979	0,74
Senador Camara	-276.300	0,75
Jardim Sulacap	137.270	0,75

Variável	beta estimado	p-valor
Anil	45.169	0,76
Cidade Nova	-108.211	0,76
Maria da Graça	-90.662	0,77
Tomas Coelho	71.113	0,77
Tanque	-70.940	0,77
Cordovil	109674	0,78
Benfica	-169.702	0,78
Parada de Lucas	93.023	0,79
Inhauma	-58.041	0,80
Cavalcanti	-148.823	0,81
Quintino Bocaiuva	-43.892	0,81
Taquara	-17.826	0,81
Abolição	62.093	0,82
Guaratiba	137.508	0,82
Bras de Pina	181.284	0,83
Vaz Lobo	-51.965	0,84
Santa Teresa	36.339	0,84
Portuguesa	50.321	0,85
Engenheiro Leal	155.723	0,86
Jacaré	-154.058	0,86
Pavuna	67.576	0,86
Marechal Hermes	-81.349	0,87
Guadalupe	56.654	0,87
Jardim Carioca	66.357	0,88
Pilares	-30.684	0,88
Pitangueiras	-38.333	0,88
Cosmos	-89.737	0,88

Variável	beta estimado	p-valor
Engenho de Dentro	16.577	0,88
Todos os Santos	17.433	0,88
Oswaldo Cruz	-68650	0,89
Engenho da Rainha	44.377	0,90
Vicente de Carvalho	40.269	0,90
Sampaio	32.855	0,91
Bangu	47.864	0,91
Maracana	-9.999	0,92
Cosme Velho	31.132	0,93
Curicica	-16.247	0,94
Cacuia	37.870	0,94
Rocha Miranda	-37.705	0,94
Barra de Guaratiba	63.734	0,94
Turiacu	-31.447	0,94
Encantado	-15.389	0,97
Monero	-9.447	0,97
Santa Cruz	18.040	0,98
Zumbi	-5.762	0,99
Cachambi	1.023	0,99

Fonte: elaborado pelo autor