FORMAÇÃO DE PREÇOS DOS IMÓVEIS DE PORTO ALEGRE: UMA ANÁLISE VIA MODELO HEDÔNICO[1]

REAL ESTATE PRICING IN PORTO ALEGRE: A HEDONIC MODEL ANALISYS

Lorenzzo Soares Santos[2]

RESUMO

O presente artigo fornece estimativas da influência das variáveis estruturais, amenidades e localização na formação dos preços de venda e aluguel dos imóveis residenciais ofertados na cidade de Porto Alegre. Utilizando base de dados secundária de 101.016 anúncios extraídos de portais imobiliários em janeiro de 2026, a pesquisa estimou modelos de preços hedônicos via regressão linear múltipla. As evidências obtidas indicam área e localização como os atributos mais fortemente associados aos preços em ambos os mercados, com variações na localização gerando amplitude de 196% no preço de venda e de 306% no aluguel entre bairros distintos. A variável suítes operou como proxy de qualidade construtiva, com prêmios de 26,7% em vendas e 23,8% em aluguéis. A comparação entre mercados sugere que compradores são mais sensíveis à metragem, enquanto locatários atribuem maior peso a amenidades de uso imediato. A hierarquia espacial de bairros mostrou-se estável na comparação com os resultados de estudos anteriores sobre Porto Alegre.

Palavras-chave: Mercado Imobiliário. Preços Hedônicos. Economia Urbana.

ABSTRACT

This article provides estimates of the influence of structural variables, amenities, and location on the price formation of residential properties listed for sale and rent in the city of Porto Alegre. Using a secondary dataset of 101,016 listings extracted from real estate portals in January 2026, the research estimated hedonic pricing models via multiple linear regression. The evidence obtained indicates floor area and location as the attributes most strongly associated with prices in both markets, with location variations generating a spread of 196% in sale prices and 306% in rents across different neighborhoods. The number of en-suite bedrooms operated as a proxy for construction quality, with associated premiums of 26.7% in sales and 23.8% in rentals. The comparison between markets suggests that buyers are more sensitive to floor area, while tenants assign greater weight to amenities of immediate use. The spatial hierarchy of neighborhoods proved stable when compared to the results of previous studies on Porto Alegre.

Keywords: Real Estate. Hedonic Prices. Urban Economics.

1 INTRODUÇÃO

A habitação constitui um dos componentes da economia urbana contemporânea. Segundo a Pesquisa de Orçamentos Familiares (IBGE, 2019), os gastos com moradia representam aproximadamente 23,9% da despesa total das famílias brasileiras. Conforme

[1]Trabalho de Conclusão de Curso apresentado, em 2026/1, ao Departamento de Economia e Relações Internacionais da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Bacharel(a) em Ciências Econômicas, sob orientação do Prof. Dr. Sabino da Silva Porto Junior.

[2]Graduando(a) em Ciências Econômicas pela Universidade Federal do Rio Grande do Sul. E-mail: lorenzzo.soares02@gmail.com.

Balchin e Kieve (1986), a habitação é simultaneamente bem de consumo, reserva de valor e instrumento de proteção contra inflação.

A formação dos preços imobiliários nas grandes cidades brasileiras resulta da interação entre atributos estruturais do imóvel, características do entorno urbano e condições macroeconômicas de crédito e renda (Hermann; Haddad; 2005). Na cidade de Porto Alegre, essa dinâmica ganha relevância em razão das transformações recentes do tecido urbano, advindas das enchentes de maio de 2024.

De acordo com o Índice FipeZap[3], o preço médio do metro quadrado em Porto Alegre atingiu R$ 7.579 em abril de 2026, com variação acumulada de 4,48% nos doze meses anteriores (FipeZap, 2026). No biênio 2024-2025, a capital registrou valorizações anuais de 7,73% e 6,52% respectivamente — as maiores em uma década (FipeZap, 2025; 2026). Esse movimento ocorre em um contexto de declínio populacional: O Censo 2022 registrou redução de 3,2% na população de Porto Alegre em relação a 2010 (IBGE, 2023), sugerindo que a valorização imobiliária pode não ser impulsionada somente por crescimento demográfico.

O ordenamento territorial de Porto Alegre é regido, desde 2000, pelo Plano Diretor de Desenvolvimento Urbano Ambiental (PDDUA), que organiza o território em macrozonas com parâmetros próprios de uso do solo, densidade e índices construtivos (Porto Alegre, 1999; 2010). Em 2026, contudo, a cidade aprovou um novo marco urbanístico — o Plano Diretor Urbano Sustentável (PDUS) —, que substituirá o PDDUA e tende a reconfigurar os padrões de oferta e verticalização nos próximos anos (Porto Alegre, 2026). O presente estudo capta, portanto, a estrutura de preços vigente imediatamente antes dessa transição regulatória.

Para compreender a formação de preços de imóveis, uma das abordagens utilizadas é o modelo de preços hedônicos. Formalizado por Rosen (1974), a teoria parte do pressuposto que um imóvel é um bem heterogêneo — sendo composto por múltiplas características que variam com base na unidade, diferente de bens homogêneos como commodities, cujas unidades são intercambiáveis. Na perspectiva de Lancaster (1966), o preço de um imóvel não reflete um valor unitário, mas a soma dos preços implícitos de cada uma de suas características físicas e locacionais, tanto observáveis diretamente (área, dormitórios) quanto de natureza indireta, que embora não sejam transacionadas em mercados próprios, são capitalizadas no preço final do imóvel; isto é, o mercado atribui valor a essas características por meio do diferencial de preço entre imóveis que possuem ou não essas características (Harrison; Rubinfeld, 1978).

A aplicação empírica dos preços hedônicos difundiu-se a partir do final dos anos 1970. Harrison e Rubinfeld (1978) utilizaram o arcabouço hedônico para investigar como atributos não diretamente transacionáveis — como qualidade do ar e criminalidade — são capitalizados nos preços de imóveis da região metropolitana de Boston, evidenciando que compradores pagam, implicitamente, por características do entorno que não estão presentes nos contratos de compra e venda. Em Porto Alegre, González (1993) transpôs essa abordagem para o mercado e aluguéis.

O objetivo geral deste trabalho é analisar os determinantes dos preços de imóveis residenciais da cidade de Porto Alegre em janeiro de 2026, para os mercados de venda e locação, aplicando a metodologia de preços hedônicos. Como objetivos específicos, o trabalho busca revisitar e atualizar as hipóteses de González (1993), transpondo-as para uma base de amostras digitais; estimar modelos de regressão múltipla, incorporando variáveis de amenidades como proxies de qualidade; submeter o modelo a uma bateria de diagnósticos

[3] O Índice FipeZAP é um indicador que acompanha a variação dos preços dos imóveis nas principais cidades do Brasil. Desenvolvido e calculado pela Fundação Instituto de Pesquisas Econômicas (Fipe), sua metodologia baseia-se na agregação e no processamento de dados secundários extraídos de anúncios online de venda e locação de imóveis.

econométricos; e comparar os padrões de precificação entre os mercados de venda e aluguel, identificando quais atributos são mais valorizados em cada segmento.

A contribuição do artigo reside em duas dimensões. A primeira é a atualização do trabalho de González (1993): partindo de uma escala amostral de 504 observações coletadas presencialmente, o presente trabalho mobiliza uma de 68 mil, extraídas de portais de anúncios online. A segunda é de natureza metodológica: a incorporação de variáveis de amenidades (piscina, churrasqueira, elevador, suítes) como substitutos observáveis dos índices de qualidade coletados em vistoria por González.

O artigo está estruturado em cinco seções, incluindo a introdução. A seção 2 apresenta o referencial teórico, abrangendo a formalização do modelo de Rosen (1974) e aplicações empíricas, e a discussão sobre formas funcionais e diagnósticos econométricos. A Seção 3 detalha os procedimentos metodológicos: origem e tratamento dos dados, classificação dos bairros, definição de variáveis e especificação dos modelos. A Seção 4 reporta e integra os resultados de venda e aluguel, comparando-os entre si e com os estimados por González (1993). A seção 5 sintetiza as considerações finais.

Em conformidade com as Recomendações para utilização da Inteligência Artificial Generativa (IAG) para os trabalhos acadêmicos no Curso de Ciências Econômicas, elaboradas pelo Núcleo Docente Estruturante do Curso de Ciências Econômicas (NDE/ECO) em agosto de 2025, declaro que utilizei ferramentas de Inteligência Artificial Generativa na elaboração deste trabalho: o ChatGPT (OpenAI) na busca inicial de referências teóricas e o Claude (Anthropic), versões Opus 4.5 e 4.6, como apoio para executar o modelo de regressão, rodar os testes estatísticos em linguagem R e apoio nas correções gramaticais.

Declaro, ainda, que a elaboração do trabalho, incluindo a seleção e análise das fontes, a construção dos argumentos, a interpretação dos resultados e as conclusões apresentadas, é de minha inteira responsabilidade.

2 REFERENCIAL TEÓRICO

A teoria microeconômica trata os bens como unidades homogêneas de consumo, com seus preços sendo definidos por oferta e demanda. Lancaster (1966) rompeu com essa tradição ao propor que os consumidores não derivam utilidade dos bens em si, mas das características ou propriedades que esses bens possuem. Essa formulação abriu caminho para a análise de bens diferenciados — como imóveis, automóveis e vinhos — cujo preço reflete um conjunto de atributos heterogêneos.

2.1 MODELO HEDÔNICO

Rosen (1974) formalizou a teoria dos preços hedônicos, detalhando como o preço de um bem composto pode ser decomposto em valores implícitos atribuíveis a cada uma de suas características. Conforme Rosen (1974, p. 34), o preço de mercado de um bem diferenciado constitui uma "função hedônica" que relaciona o preço observado ao vetor de seus atributos, para obter os preços-sombra de cada característica do produto. Os coeficientes da equação hedônica, portanto, podem ser interpretados como preços implícitos de cada atributo do produto analisado. No caso de um imóvel residencial, o produto pode ser descrito por um vetor de n características mensuráveis, que Rosen (1974), denomina *z*.

$$z = (z_1, z_2, \ldots, z_n) \quad (1)$$

onde cada z_j representa um atributo específico do imóvel — z_1 pode representar a área útil em metros quadrados, z_2 o número de dormitórios, z_3 o número de vagas de garagem, z_4 a presença ou ausência de piscina (variável dummy), e assim por diante. Atributos locacionais, como o bairro, também integram esse vetor.

O mercado imobiliário estabelece, em equilíbrio competitivo, uma função de preços hedônicos — $P(z)$ — que relaciona o preço do imóvel ao seu vetor de atributos. Conforme formalizado por Rosen (1974, p. 38-40), essa função emerge da interação entre oferta e demanda de mercado.

Assim, do lado da demanda, cada consumidor racional escolhe uma cesta de bens que maximiza sua função de preferências, representada por:

$$U(x, z) \quad (2)$$

onde U representa a função de utilidade do consumidor; x a quantidade consumida de um bem composto numerário (que representa todos os demais gastos do consumidor além do imóvel), z o vetor de características do bem, conforme definido em (1). O consumidor está sujeito à restrição orçamentária:

$$y = x + P(z) \quad (3)$$

onde y representa a renda total do consumidor. A função de lance — que Rosen (1974) denomina θ — representa o valor máximo que o consumidor está disposto a pagar por um bem com atributos z e ainda atingir um nível de utilidade u. Com θ sendo definido por:

$$U(y - \theta, z) = u \quad (4)$$

onde $y - \theta$ representa a renda residual após o dispêndio com o bem — parcela destinada ao consumo do bem numerário x — e u representa o nível de utilidade que o consumidor obtém. A equação estabelece que, ao pagar θ pelo imóvel com características z, o consumidor destina a renda residual $y - \theta$ ao consumo dos demais bens, atingindo o nível de utilidade u.

O preço-sombra de cada atributo z_j do bem é obtido pela derivada parcial da função de lance em relação a esse atributo, que corresponde à taxa marginal de substituição entre o atributo z_j e o consumo do bem numerário x:

$$\partial\theta/\partial z_j = (\partial U/\partial z_j) / (\partial U/\partial x) \quad (5)$$

A expressão (5) representa a disposição marginal a pagar pelo atributo z_j: a redução no consumo do bem numerário x que o consumidor aceita em troca de uma unidade adicional de z_j, mantido constante o nível de utilidade (Freeman, 1979, p. 68).

Considerando o lado da oferta de mercado, cada produtor maximiza lucros escolhendo o vetor de características do produto — z — que vai definir o preço final do bem. A função de oferta $\varphi(z; \pi)$ indica o preço mínimo que o produtor aceita para ofertar um bem com atributos z, dado um nível de lucro π. Essa função reflete o custo marginal de cada atributo do produto: por exemplo, incluir uma suíte custa mais que um banheiro simples, e adicionar uma piscina ao condomínio eleva os custos fixos do empreendimento.

No equilíbrio de mercado, o preço observado $P(z)$ iguala, para cada atributo, a disposição a pagar do consumidor marginal e o custo marginal do produtor marginal:

$$\partial P/\partial z_j = \partial\theta/\partial z_j = \partial\varphi/\partial z_j \quad (6)$$

Com a derivada parcial (6) representando o preço implícito do atributo z_j; o acréscimo no preço do bem associado a uma unidade adicional dessa característica, mantidas constantes

as demais. O preço marginal dos atributos é estimado fazendo regredir o preço do bem heterogêneo sobre as quantidades de características associadas (Herman, Haddad, 2005).

$$log(P_i) = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \ldots + \beta_n z_{ni} + \varepsilon_i \quad (7)$$

onde P_i é o preço do bem i, z_{1i} a z_{ni} são os valores dos n atributos observados para esse bem, β0 é o intercepto, β1 a βn são os coeficientes a estimar, e ε_i é o termo de erro aleatório. Os coeficientes estimados $\hat{\beta}_j$ representam os preços implícitos marginais de cada atributo do imóvel. Conforme Wooldridge (2020, p. 189), na forma semilogarítmica dessa equação, o coeficiente $\hat{\beta}_j$ representa a variação percentual no preço do bem por unidade adicional do atributo z_j; na forma log-log, o coeficiente $\hat{\beta}_j$ assume interpretação de elasticidade — isto é, mede a variação percentual no preço associada a uma variação de 1% no atributo.

Cabe observar que essa primeira etapa estima a função de preços de equilíbrio do mercado, e não as funções de demanda ou oferta dos agentes (Rosen, 1974, p. 44). A identificação dessas funções estruturais requereria uma segunda etapa, que enfrenta problemas de identificação de causalidade, conforme apontado por Brown e Rosen (1982). Por essa razão, a literatura empírica concentra-se na primeira etapa (Sheppard, 1999, p. 1605).

2.2 APLICAÇÕES EMPÍRICAS

A abordagem hedônica é, segundo Sheppard (1999, p. 1595), "uma das poucas ferramentas que permitem mensurar empiricamente a valorização gerada por políticas públicas urbanas, externalidades ambientais e infraestrutura". A primeira aplicação de grande repercussão foi realizada por Harrison e Rubinfeld (1978), que utilizaram dados de 506 setores censitários de Boston para estimar a disposição a pagar por ar limpo na cidade. Para isolar o efeito da poluição — medida pela concentração de óxidos de nitrogênio (NOX) —, os autores controlaram o preço por atributos do imóvel, características da vizinhança, carga tributária e acessibilidade a centros de emprego. A contribuição central de Harrison e Rubinfeld (1978) foi indicar que atributos não transacionáveis diretamente, como qualidade do ar e criminalidade, são capitalizados nos preços dos imóveis.

Já Arraes e Sousa Filho (2008) estimaram funções hedônicas para 4.467 transações em Fortaleza entre 1995 e 2003, segmentaram a amostra por tipo de imóvel e acrescentaram, ao lado de área, idade e número de cômodos, atributos de risco e retorno voltados ao investidor. Hermann e Haddad (2005) estimaram uma equação hedônica para o aluguel residencial em São Paulo a partir de dados da POF/Fipe (1999). Sua principal contribuição metodológica foi abandonar a hipótese monocêntrica usual e testar também uma configuração duocêntrica, com dois polos de emprego — o centro histórico (Praça da Sé) e o complexo Berrini–Faria Lima —, o que altera a forma como a acessibilidade entra na função de preços. Para conter a forte multicolinearidade entre as variáveis de vizinhança, recorreram à análise fatorial, condensando-as em fatores como infraestrutura, aprazibilidade e segurança. Concluíram que a proximidade a estações de trem, a presença de áreas verdes e o zoneamento estritamente residencial elevam o aluguel, ao passo que a criminalidade o reduz, com efeito ambíguo ou não significativo para o metrô e a atividade comercial; o R^2 ajustado situou-se próximo de 0,69 nas especificações ambientais. Importa notar que, à medida que os autores incorporavam variáveis georreferenciadas ao modelo, a autocorrelação espacial dos resíduos perdia significância nos testes de Multiplicadores de Lagrange — indício direto de que a dependência espacial nasce, em boa parte, de variáveis de vizinhança omitidas.

Essa constatação introduz um segundo conjunto de estudos, voltado à dependência espacial, fenômeno que decorre da própria natureza do bem. Como imóveis vizinhos partilham amenidades de entorno que o pesquisador raramente observa por completo —

qualidade da vizinhança, paisagem, percepção de segurança —, seus preços tendem a se correlacionar no espaço; a essa fonte, ligada a variáveis omitidas, soma-se o efeito de transbordamento (spillover), já que o valor de um imóvel costuma servir de referência para a precificação dos vizinhos. Em ambos os casos viola-se a hipótese de independência entre observações do método de mínimos quadrados, e os coeficientes — sobretudo os de localização, dimensão em que a correlação opera — tornam-se enviesados ou ineficientes, com significância pouco confiável. Campos (2017), para São Paulo, mostrou que ignorar essa autocorrelação distorce os coeficientes associados à localização. Seabra et al. (2016) chegaram a resultado análogo em Recife: detectada a dependência pelo I de Moran, os autores estimaram modelos autorregressivos espaciais (SAR e SAC) que corrigiram o viés do modelo OLS e permitiram decompor os efeitos em diretos e indiretos. Nessa especificação, a vista para o mar firmou-se como a amenidade de maior impacto, com valorização direta em torno de 13% e efeito total superior a 16%.

É nos estudos sobre Porto Alegre que esta pesquisa se insere. González (1993) estimou um modelo hedônico para o aluguel de 504 apartamentos vistoriados em junho de 1992, na forma linear, com a área expressa em logaritmo para captar os retornos decrescentes da metragem:

Para o autor, a contagem de cômodos não captura a qualidade do imóvel, razão pela qual ele registrou em vistoria índices de padrão construtivo e estado de conservação. O modelo alcançou R^2 entre 0,70 e 0,85, com área e localização explicando a maior parte da variação do preço e colinearidade entre banheiros e dormitórios. González e Formoso (1995) voltaram-se a uma questão distinta: a existência, ou não, de um modelo único capaz de explicar o aluguel de qualquer apartamento da cidade. Trabalhando com a mesma amostra de 504 imóveis e empregando análise fatorial para conter a multicolinearidade, estimaram modelos para o conjunto total e para subconjuntos segmentados — por número de dormitórios e por distância ao centro. Embora o modelo geral apresentasse bom ajuste (R^2 acima de 0,80) e pudesse representar a média da cidade, as equações dos submercados diferiam de forma expressiva quanto às variáveis selecionadas e à magnitude dos coeficientes, de modo que a hipótese do modelo único não se confirmou plenamente. Entre os determinantes relevantes destacaram-se a distância aos *shopping centers* e a presença de favelas no entorno, está associada à desvalorização do imóvel.

Ainda para Porto Alegre, trabalhos recentes ampliaram o escopo metodológico dos estudos. Julio (2021) utilizou uma abordagem hierárquico-espacial para modelar os preços de apartamentos construídos na cidade entre 2007 e 2016, incluindo variáveis intrínsecas (área, dormitórios, vagas) e extrínsecas (amenidades urbanas do entorno, infraestrutura do bairro). Os resultados indicaram concentração de amenidades urbanas positivas no entorno da região central, com imóveis de padrão mais alto gravitando em direção ao Centro — embora o grande número de edifícios em bairros mais distantes tenha reduzido a sensibilidade dos consumidores a essas amenidades. Vecchia (2025) estimou modelos preditivos para imóveis comerciais em Porto Alegre, comparando regressão múltipla, modelagem multinível e econometria espacial. O modelo espacial apresentou desempenho superior nas métricas de erro de previsão (MAE e RMSE), e variáveis como rendimento nominal e densidade populacional mostraram-se significativas em todas as especificações (Vecchia, 2025).

3 PROCEDIMENTOS METODOLÓGICOS

Toma-se como referência para a estimação a metodologia desenvolvida por González (1993). Em sua dissertação, o autor reuniu 504 apartamentos ofertados para locação em Porto Alegre em junho de 1992 — identificados junto a imobiliárias e a anúncios de jornal —, registrando o aluguel pedido e um conjunto de características obtidas por vistoria externa

(padrão de acabamento, idade aparente, material da fachada e condições do entorno), complementadas pelo cadastro da Secretaria Municipal da Fazenda (área total e idade). A função de preços reunia atributos estruturais (área, dormitórios, banheiros e garagem), de qualidade (padrão construtivo e estado de conservação) e de localização (distância ao centro e a polos comerciais, além de uma classificação dos bairros por renda média).

Os modelos foram estimados por mínimos quadrados ordinários, na forma linear clássica. Diante da forte colinearidade entre as características do imóvel, González recorreu à análise fatorial para agrupar as variáveis correlacionadas — admitindo apenas uma representante de cada fator na equação — e selecionou o conjunto final pelo procedimento stepwise. A forma funcional foi definida empiricamente: o autor constatou que logaritmizar o valor do aluguel não trazia ganho relevante, ao passo que expressar a área em logaritmo apresentava melhor comportamento do que a forma em nível, captando os retornos decrescentes da metragem. A adequação foi aferida por um conjunto de cuidados econométricos: testes t para os regressores e F para a regressão, coeficiente de determinação e análise de resíduos abrangendo média, heterocedasticidade, autocorrelação (teste de Durbin-Watson), normalidade e identificação de outliers, com os limites de ajuste tomados da norma de avaliação de imóveis urbanos da ABNT (NB-502).

O presente trabalho revisita as hipóteses levantadas por esse autor, transpondo-as para a realidade urbana de 2026. A atualização metodológica reside na transição de uma amostragem presencial para o uso de dados imobiliários disponíveis em portais de anúncio online. Adotou-se uma abordagem quantitativa baseada em dados secundários de corte transversal, com estimação de modelos de regressão múltipla pelo método dos Mínimos Quadrados Ordinários (MQO).

3.1 BASE DE DADOS

A base de dados utilizada para a estimação dos modelos origina-se da plataforma digital de anúncio de imóveis ZAP Imóveis, extraída via web scraping[4] no período de janeiro de 2026, resultando em uma amostra bruta de 101.016 observações para vendas e 12.105 para aluguéis.

Por se tratar de uma base não oficial, em que imóveis podem ser anunciados por qualquer indivíduo, foi realizado um processo de limpeza em cinco etapas: a exclusão de imóveis em municípios da região metropolitana (Canoas, Gravataí, Novo Hamburgo e São Leopoldo); a remoção de terrenos e imóveis de tipologia exclusivamente comercial; a remoção de unidades com valores faltantes (NA) ou ilógicos, como área e preço zerados ou negativos; a remoção de duplicatas via cruzamento de coordenadas geográficas, área útil, número de quartos e preço; e a remoção de valores extremos pelo método de percentil (cortes no 1º e 99º percentis de preço e área).

Após esse processo, a amostra de vendas foi reduzida a 70.986 observações, e a de aluguéis a 8.737 unidades. Uma etapa adicional de identificação de observações influentes pela Distância de Cook foi aplicada após a estimação dos modelos, conforme descrito na Seção 3.3.4.

3.2 CLASSIFICAÇÃO DOS BAIRROS

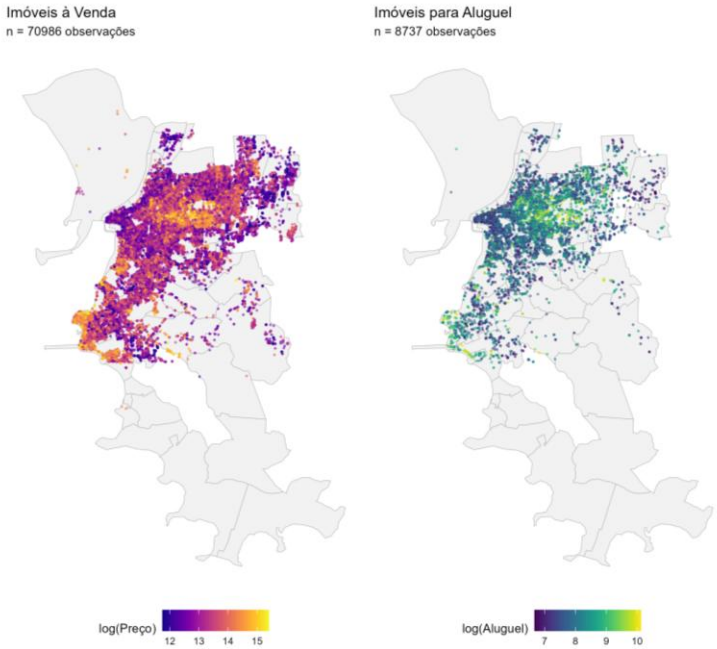
Um dos desafios na comparação de modelos com intervalos temporais amplos é a mutação do tecido urbano. No trabalho de González (1993), o autor precisou excluir regiões

[4] *Web scraping* é a extração automatizada de dados de páginas da internet por meio de scripts de programação.

devido à baixa amostragem e à inexistência oficial de delimitações de novos bairros. No presente estudo, a geolocalização dos anúncios permitiu que a classificação por bairro seguisse o mapa oficial de zoneamento de Porto Alegre.

Novos bairros criados por legislações municipais recentes e desmembramentos de regiões anteriores foram capturados pela geolocalização dos anúncios.

Figura 1 - Distribuição das amostras



Fonte: Elaboração Própria com mapas municipais (IBGE)

3.3 ESPECIFICAÇÃO DO MODELO

A teoria dos preços hedônicos, conforme formalizada por Rosen (1974), não prescreve uma forma funcional específica para a equação de equilíbrio. A literatura empírica explora três alternativas: a forma linear, a semilogarítmica (log-linear) e a dupla-logarítmica (log-log), com hipóteses distintas sobre a relação entre preço e atributos (Wooldridge, 2020, cap. 6).

Na forma linear, o preço Pi é modelado como função aditiva das características, e cada coeficiente β representa o acréscimo absoluto associado a uma unidade adicional do atributo. Na forma semilogarítmica, aplica-se o logaritmo natural ao preço, e os coeficientes passam a representar variações percentuais no preço do imóvel. Na forma log-log, tanto o preço quanto uma ou mais variáveis explicativas são transformadas em logaritmo, e o coeficiente assume interpretação de elasticidade.

González (1993) emprega transformações logarítmicas para capturar retornos decrescentes da metragem. Seguindo essa prescrição, o presente estudo adotou uma especificação mista, na qual a variável área entra em logaritmo (forma log-log) e as demais variáveis explicativas entram em nível (forma log-linear), tendo o logaritmo do preço como variável dependente. A adaptação do arcabouço de González (1993) a dados de portais online impõe ainda uma restrição: as variáveis de padrão construtivo (PC) e estado de conservação (EC), que o autor coletava em vistoria, não existem em bases de anúncio. Como contorno, foram incorporadas amenidades observáveis — suítes, piscina, churrasqueira e elevador — como proxies de qualidade, distinguindo-se banheiros comuns de suítes e utilizando-se dummies de bairro como efeitos fixos locacionais. Optou-se por não incluir a variável banheiros de forma isolada, que, estimada juntamente com suítes, apresentava

multicolinearidade elevada, uma vez que uma suíte é, por definição, um dormitório com banheiro. A especificação estimada é:

$$\log(P_i) = \beta_0 + \beta_1 \cdot log(At_i) + \beta_2 \cdot D_i + \beta_3 \cdot V_i + \beta_4 \cdot S_i + \beta_5 \cdot Sp_i + \beta_6 \cdot Ch_i + \beta_7 \cdot E_i + \Sigma\, \delta_j \cdot Ti_j + \Sigma\, \gamma_k \cdot Ba_k + \varepsilon_i \quad (8)$$

em que P_i representa o preço do imóvel i; At, a área útil em m²; D, o número de dormitórios; V, o número de vagas de garagem (presente apenas no modelo de vendas); S, o número de suítes; Sp, Ch e E, as variáveis binárias de piscina, churrasqueira e elevador, respectivamente; Ti e Ba, os vetores de dummies de tipologia e bairro; e ε_i, o termo de erro. Como a área entra em logaritmo, o coeficiente β_1 associado a ela assume a interpretação de elasticidade:

$$\partial log(P)/\partial log(At) = \beta_1 \quad (9)$$

de modo que um aumento de 1% na área útil está associado a uma variação de β_1% no preço. Quando $\beta_1 < 1$, tem-se retornos decrescentes da metragem, González (1993).

Para variáveis binárias em equações semilogarítmicas, a interpretação do efeito percentual requer a correção proposta por Halvorsen e Palmquist (1980). No presente modelo, essa correção é materialmente relevante para suítes (c = 0,237), onde a aproximação linear indicaria prêmio de 23,7%, mas o efeito exato, dado por $(e^c - 1) \times 100$, é de 26,7%. Todos os efeitos percentuais reportados nas tabelas de resultados foram calculados pela fórmula exata.

Para a seleção entre os três modelos, utilizou-se o Critério de Informação de Akaike (Akaike, 1974):

$$AIC = -2\, ln(\hat{L}) + 2k \quad (10)$$

onde $\hat{L}$ é o valor maximizado da função de verossimilhança e k o número de parâmetros. O AIC penaliza modelos com mais parâmetros pelo termo 2k, de modo que um AIC menor indica melhor equilíbrio entre ajuste e parcimônia. O R^2 ajustado foi utilizado como critério complementar.

3.4 DIAGNÓSTICOS ECONOMÉTRICOS

Após a estimação dos modelos, os pressupostos da regressão por MQO foram avaliados por quatro diagnósticos: multicolinearidade entre regressores, heterocedasticidade dos erros com adoção de erros-padrão robustos, identificação de observações influentes por Distância de Cook e verificação da normalidade dos resíduos.

O diagnóstico de multicolinearidade foi realizado pelo Fator de Inflação da Variância (VIF), definido como (Wooldridge, 2020, p. 86):

$$VIF_j = 1 / (1 - R^2_j) \quad (11)$$

em que R^2_j é obtido pela regressão auxiliar da variável x_j sobre as demais covariáveis. No contexto do presente modelo, o VIF foi aplicado para avaliar a sobreposição entre suítes e banheiros — variáveis que, por construção, descrevem a mesma dimensão do imóvel. González (1993, p. 89) já reportava colinearidade entre essas variáveis em sua amostra de aluguéis. O limiar de 10, adotado pela literatura como indicativo de problema (Belsley, Kuh e Welsch, 1980), não foi atingido por nenhuma variável: todos os VIFs situaram-se abaixo de 2,6, e a exclusão de banheiros em favor de suítes eliminou a fonte de redundância.

A presença de heterocedasticidade foi avaliada pelo teste de Breusch e Pagan (1979), cuja lógica consiste em verificar se a dispersão dos resíduos varia com os regressores. A estatística do teste é:

$$BP = n \cdot R^2_{aux} \quad (12)$$

em que n é o número de observações e R^2_{aux} é o R^2 da regressão auxiliar dos resíduos ao quadrado sobre os regressores originais; sob a hipótese nula de variância constante, a estatística distribui-se como uma qui-quadrado com k graus de liberdade. Em dados de corte transversal com unidades de tamanhos muito distintos, a heterocedasticidade é frequente (Wooldridge, 2020, cap. 8): na presente base, em que coexistem imóveis de 25 m² e de 500 m², a rejeição da hipótese nula era esperada.

A correção adotada foi o estimador HC1, proposto por MacKinnon e White (1985), que ajusta a matriz de covariância robusta de White (1980) por um fator de escala:

$$\hat{V}HC1 = (n/(n-k)) \cdot (X'X)^{-1} \cdot X' \, diag(\hat{e}_i^2) \cdot X \cdot (X'X)^{-1} \quad (13)$$

correção relevante quando o número de parâmetros é elevado em relação à amostra — caso do presente modelo, com cerca de 100 regressores (incluindo 83 dummies de bairro). Os coeficientes estimados não se alteram com a adoção de erros robustos; o que muda são as variâncias estimadas e, com elas, os erros-padrão, as estatísticas t e os p-valores.

Como a base é composta por anúncios de portais online, o modelo está sujeito a observações atípicas — anúncios com preços incorretos, combinações incomuns de atributos ou erros de preenchimento. Para identificá-las, utilizou-se a Distância de Cook (Cook, 1977), que sintetiza a influência global de cada observação sobre o vetor de coeficientes estimado:

$$D_i = (\hat{\beta} - \hat{\beta}_{(i)})' \cdot (X'X) \cdot (\hat{\beta} - \hat{\beta}_{(i)}) / (k \cdot s^2) \quad (14)$$

onde $\hat{\beta}(i)$ é o vetor estimado sem a observação i, k é o número de parâmetros e s^2 é o quadrado médio dos resíduos. Essa medida tem a forma computacional equivalente:

$$D_i = (\hat{e}_i^2 / (k \cdot s^2)) \cdot (h_{ii} / (1 - h_{ii})^2) \quad (15)$$

em que $\hat{e}_i$ é o resíduo da observação i e h_{ii}, o i-ésimo elemento da diagonal da matriz hat, que mede a alavancagem — a distância da observação ao centroide dos regressores. Observações com D_i acima do limiar convencional de 4/n foram sinalizadas como influentes e removidas antes da estimação final.

O pressuposto de normalidade dos resíduos é necessário para a validade exata dos testes t e F em amostras finitas. Para verificá-lo, aplicou-se o teste de Shapiro-Wilk (Shapiro; Wilk, 1965) sobre subamostra aleatória de 5.000 observações. A estatística W é calculada como:

$$W = (\Sigma \, a_i \cdot x_{(i)})^2 / \Sigma (x_i - \bar{x})^2 \quad (16)$$

em que $x_{(i)}$ são as estatísticas de ordem da amostra e a_i, coeficientes tabelados; valores de W próximos de 1 indicam aderência à normalidade. Contudo, em amostras da magnitude utilizada (n > 68.000), o teste tende a rejeitar a hipótese nula mesmo diante de desvios irrelevantes na prática. Como o Teorema do Limite Central garante que a distribuição dos estimadores de MQO é aproximadamente normal, independentemente da distribuição dos erros, desde que os demais pressupostos estejam razoavelmente satisfeitos (Wooldridge, 2020, cap. 4), a avaliação baseou-se prioritariamente na inspeção dos gráficos Q-Q e do histograma de resíduos (Seção 4).

4 ANÁLISE DOS RESULTADOS

Após a limpeza da base e a remoção de observações influentes pela Distância de Cook, a amostra final compreendeu 68.051 imóveis para o modelo de vendas e 8.338 para o de aluguel. Foram estimadas três especificações progressivas para cada mercado, conforme

descrito na Seção 3.3, e a seleção do modelo final baseou-se no Critério de Informação de Akaike e no R^2 ajustado. A Seção 4.1 apresenta os resultados para vendas, a Seção 4.2 para aluguéis, e a Seção 4.3 sintetiza a comparação entre os dois mercados.

4.1 RESULTADOS DO MODELO DE VENDA

Os coeficientes reportados nas tabelas a seguir referem-se ao Modelo 3 (log-log em área) após remoção por Cook, com erros-padrão robustos HC1. Cabe ressalvar que, por se tratar de dados de corte transversal e de natureza observacional, os coeficientes expressam correlações parciais entre atributos e preços, e não relações de causalidade (Wooldridge, 2020, cap. 1).

Tabela 1 - Especificações do modelo de vendas

Especificação	*R^2 ajustado*	AIC	N
Modelo 1	0,665	80.888	70.986
Modelo 2	0,761	57.028	70.986
Modelo 3	0,792	46.991	70.986
Modelo final	0,836	26.722	68.051

Fonte: Elaboração Própria (2026)

A inclusão de amenidades e suítes no Modelo 2 elevou o R^2 ajustado de 0,665 para 0,761 e reduziu o AIC em quase 24 mil pontos — um ganho atribuível a variáveis que captam qualidade construtiva, em linha com a ênfase de González (1993, p. 48) na insuficiência da contagem de cômodos. A transformação logarítmica da área no Modelo 3 produziu ganho adicional (R^2 = 0,792), e a remoção por Cook elevou o R^2 final a 0,836, com estatística F = 3.660,09 ($p<0,001$).

Tabela 2 – Coeficientes das variáveis independentes do modelo de vendas

Variável	Coef	Erro padrão	t_valor	Sig
(Intercept)	9,8787	0,1948	50,7174	***
log(area)	0,5773	0,0037	156,762	***
Dormitórios	-0,0004	0,0018	-0,2156	NS
Vagas	0,0846	0,0012	68,1170	***
Suítes	0,2367	0,0018	129,962	***
Piscina	0,1406	0,0034	41,1478	***
Churrasqueira	0,0610	0,0027	22,3397	***
Elevador	0,1382	0,0029	47,7633	***

Fonte: Elaboração Própria (2026)

O coeficiente de log(área), de 0,577, indica que um aumento de 1% na área útil está associado a um aumento de 0,58% no preço de venda, mantidas constantes as demais variáveis. Por se tratar de elasticidade inferior à unidade, o resultado sugere retornos decrescentes da metragem — padrão documentado por González (1993) para aluguéis em Porto Alegre e Sheppard (1999) em contextos internacionais.

A variável dormitórios apresentou coeficiente não significativo (−0,000; $p > 0,10$). Na presença de log(área), o efeito de dimensionamento é absorvido pela metragem: imóveis com mais quartos são, quase invariavelmente, imóveis maiores. No Modelo 2 (área em nível), dormitórios era significativo ($\beta = 0,150$; $p < 0,001$), o que indica que a perda de significância decorre da mudança de forma funcional. González (1993, p. 89) já reportava colinearidade entre dormitórios e área.

Suítes apresentou a maior correlação parcial dentre os atributos estruturais: coeficiente de 0,237, associado a um acréscimo de 26,7% por unidade adicional ($e^{0,237} - 1 = 0,267$). Enquanto González utilizava índices subjetivos de padrão construtivo (*PC*) e estado de conservação (*EC*) coletados em vistoria, o presente resultado sugere que suítes funciona como proxy observável dessa mesma dimensão qualitativa (González, 1993, p. 48).

Vagas de garagem apresentaram coeficiente de 0,085, associado a acréscimo de 8,8% por vaga. As dummies de amenidades apresentaram correlações positivas e significativas: piscina (+15,1%), elevador (+14,8%) e churrasqueira (+6,3%). Essas amenidades operam como sinalizadores de padrão construtivo — papel análogo ao das variáveis PC e EC de González (1993). A churrasqueira, disseminada no Rio Grande do Sul, apresentou menor efeito diferenciador.

Tabela 3 – Coeficientes de localização do modelo de vendas: maiores prêmios e descontos

Variável	Coef	Std.err	t_valor	Sig
J.Europa	0,6460	0,1708	3,7836	***
Três Fig.	0,5514	0,1702	3,2391	**
Bela Vista	0,4831	0,1701	2,8391	**
Moinhos de V.	0,4715	0,1701	2,7716	**
S.R de Lima	-0,4414	0,1718	-2,5697	*
Rubem Berta	-0,4229	0,1710	-2,4727	*
J.Leopoldina	-0,3566	0,1705	-2,0913	*
Pq. St. Fé	-0,3436	0,1706	-2,0136	*

Fonte: Elaboração própria. Tabela completa no Apêndice B

A amplitude entre Jardim Europa (+0,646) e Santa Rosa de Lima (−0,441) corresponde a uma diferença de aproximadamente 196% no preço de dois imóveis com atributos físicos idênticos. Os bairros com maiores prêmios — Jardim Europa, Três Figueiras, Bela Vista, Moinhos de Vento — situam-se na Macrozona 1 (Cidade Radiocêntrica) do PDDUA, a área com maior infraestrutura e índices construtivos de Porto Alegre. Os bairros com maiores descontos — Rubem Berta, Santa Rosa de Lima, Jardim Leopoldina — situam-se nas Macrozonas 3 e 6 (Cidade Xadrez e Eixo Lomba-Restinga), de ocupação mais recente e menor oferta de serviços.

Tabela 4 – Comparação de resultados e posições

Bairro	Posição 1993	Coef	Posição 2026
Bela Vista	1º	0,483**	3º
Farroupilha	2º	0,31	~15º
Rio Branco	3º	0,392*	7º
Centro	7º	0,015	~40º
Rubem Berta	16º–17º	−0,423*	~80º
Jardim do Salso	18º	0,166	~25º

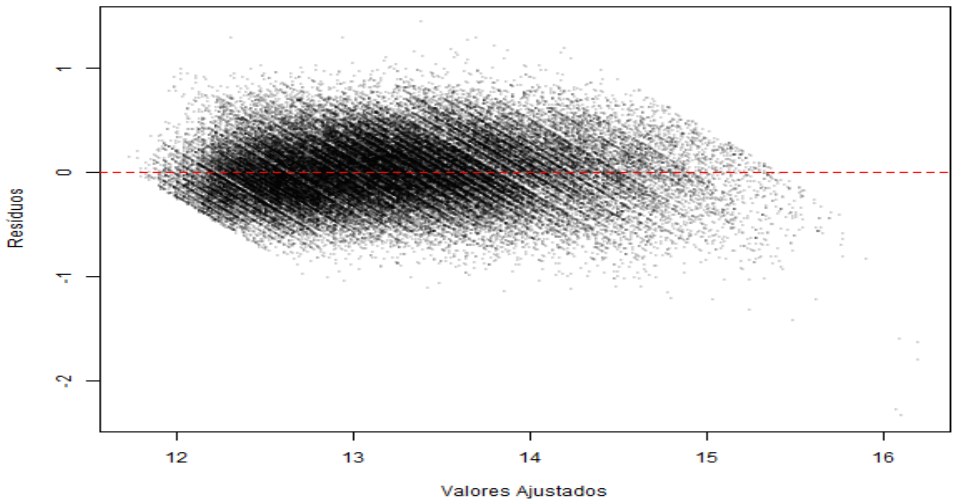
Fonte: Elaboração Própria. Índice de González: Tabela 5.12 (González, 1993, p. 50)

A comparação com González (1993, p. 50, Tabela 5.12) — que utilizava um índice contínuo de 0 a 100 baseado na renda do Censo de 1980 — revela estabilidade na hierarquia locacional ao longo de três décadas. Bela Vista (índice 100 em 1993) permanece entre os maiores prêmios em 2026. Rubem Berta (índice 5) permanece entre os maiores descontos. O Centro Histórico (índice 29) apresenta coeficiente próximo de zero em 2026 (+0,015, não significativo), o que é coerente com o esvaziamento residencial observado em centros de

capitais brasileiras. Jardim do Salso (índice 0) passou a apresentar coeficiente positivo (+0,166), refletindo a expansão de empreendimentos na zona leste. Jardim Europa e Três Figueiras, que lideram o ranking de 2026, não figuravam na amostra de González.

Os VIFs situaram-se abaixo de 2,6 (Belsley, Kuh e Welsch, 1980). O teste de Breusch-Pagan resultou em BP = 16.326,69 ($p < 0,001$), rejeitando a hipótese nula de variância constante. Em dados de corte transversal com unidades de tamanhos distintos, a presença de heterocedasticidade é frequente (Wooldridge, 2020, cap. 8). Na presente base, onde coexistem imóveis de 25 m² e de 500 m², a rejeição era esperada. Os erros robustos HC1 (MacKinnon; White, 1985) foram adotados para toda a inferência.

Figura 2 - Resíduos vs Valores Ajustados (vendas)

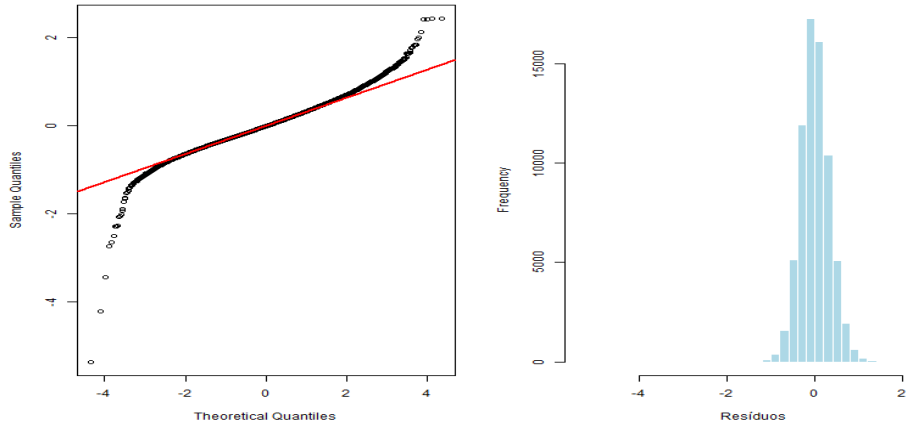


Fonte: Elaboração Própria (2026)

A Figura 2 apresenta os resíduos plotados contra os valores ajustados. Observa-se concentração de pontos em torno da linha zero, com dispersão homogênea na faixa entre 12 e 15. Na faixa inferior (abaixo de 12,5), nota-se leve aumento da dispersão — um padrão de funil coerente com a heterocedasticidade detectada pelo teste de Breusch-Pagan.

O teste de Shapiro-Wilk (Shapiro; Wilk, 1965), aplicado sobre subamostra de 5.000 observações, resultou em W = 0,996 ($p < 0,001$). A rejeição de H_0 é esperada em amostras dessa magnitude. O Teorema do Limite Central garante a validade assintótica dos estimadores (Wooldridge, 2020, cap. 4).

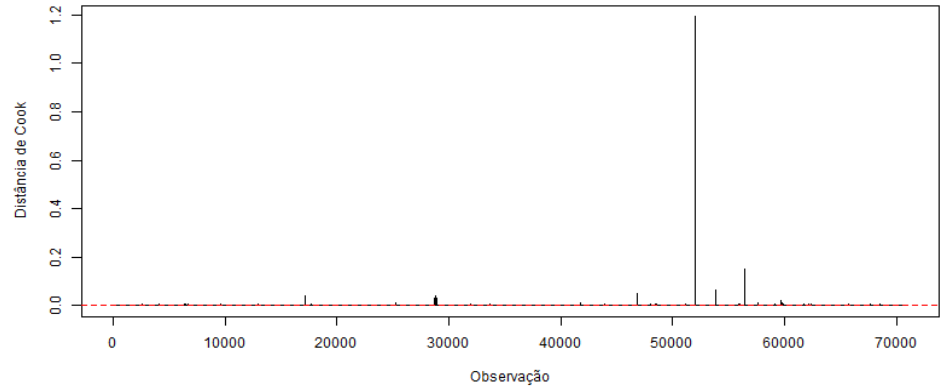
Figura 3 – Q-Q Plot e Histograma dos Resíduos do modelo de venda



Fonte: Elaboração Própria (2026)

Na Figura 3, o gráfico Q-Q mostra que os resíduos acompanham a reta teórica na região central (quantis −2 a +2), com desvios nas caudas. O histograma exibe distribuição aproximadamente simétrica em torno de zero, com leve assimetria à esquerda. Dada a dimensão amostral (n > 68.000), esses desvios não comprometem a inferência.

Figura 4- Distância de Cook do modelo de vendas



Fonte: Elaboração Própria (2026)

Na Figura 4, a maioria das observações situa-se abaixo do limiar $4/n = 5{,}63 \times 10^{-5}$. Poucos picos isolados se destacam — o mais acentuado próximo à observação 54.000, com D > 1,0. A remoção por Cook eliminou aproximadamente 4% da amostra (de 70.986 para 68.051) e esteve associada a elevação do R^2 de 0,792 para 0,836.

4.2 RESULTADOS DO MODELO DE ALUGUEL

No modelo de aluguel, a elasticidade-área foi de 0,406 — inferior à de vendas (0,577). A duplicação da área está associada a um aumento de ~32% no aluguel ($20^{0{,}406-1} \approx 0{,}32$), contra ~49% no preço de compra. Esse contraste sugere que locatários são menos sensíveis à metragem que compradores, possivelmente porque a decisão de aluguel é de horizonte mais curto e o custo mensal pesa de forma direta.

Tabela 5 – Coeficientes das variáveis independentes do modelo de aluguel

Variável	Coef	Std.err	t_valor	Sig
(Intercept)	6,0352	0,1526	39,5399	***
log(area)	0,4057	0,0111	36,4622	***
Dormitórios	0,0362	0,0062	5,8867	***
Suítes	0,2132	0,0060	35,5449	***
Piscina	0,1435	0,0108	13,3327	***
Churrasqueira	0,1326	0,0087	15,1801	***
Elevador	0,0862	0,0087	9,9268	***

Fonte: Elaboração Própria (2026)

Diferentemente do modelo de vendas, dormitórios apresentou coeficiente significativo (0,036; p < 0,01), associado a acréscimo de 3,7% por unidade. No mercado de locação, a configuração funcional pode operar como critério de busca — locatários filtram por "3 dormitórios", não por "90 m²" —, o que confere relevância ao número de quartos que, no modelo de vendas, é absorvida por log(área). González (1993, p. 89) reportava colinearidade entre dormitórios e área, mas utilizava QUARTOS em parte de suas especificações, obtendo coeficientes significativos.

Suítes mantiveram prêmio associado de 23,8% — próximo ao de vendas (26,7%). A churrasqueira apresentou coeficiente de 0,133, mais que o dobro do observado em vendas (0,061). Para locatários, a disponibilidade imediata de áreas de lazer pode ter maior peso, dado que o aluguel não envolve expectativa de reformas. O elevador apresentou coeficiente menor que em vendas (0,086 vs 0,138).

Tabela 6 – Coeficientes de localização do modelo de aluguel: maiores prêmios e descontos

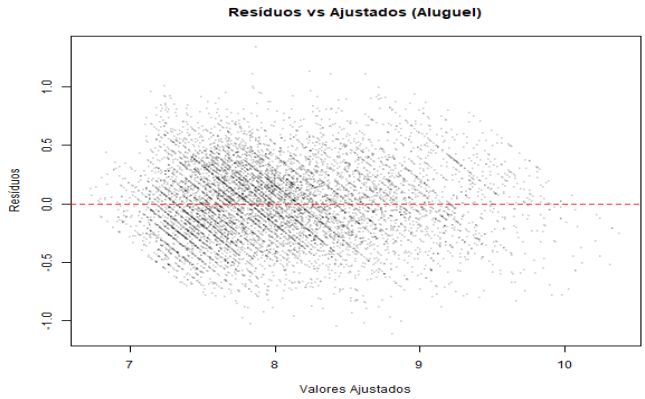
Variável	Coef	Std.err	t_valor	Sig
Jardim Europa	0,5482	0,1524	3,597	***
Três Figueiras	0,382	0,1487	2,5684	*
Pedra Redonda	0,3392	0,1574	2,1544	*
Montserrat	0,3156	0,1484	2,1265	*
Belem Velho	-0,8543	0,1917	-4,4563	***
Rubem Berta	-0,8182	0,1585	-5,1606	***
Parque Santa Fé	-0,6610	0,1619	-4,0842	***
Jardim Leopoldina	-0,6463	0,1592	-4,0586	***

Fonte: Elaboração Própria. Tabela completa no Apêndice C (2026)

O padrão espacial é similar ao de vendas. A comparação com González (1993) é aqui pertinente, uma vez que seu modelo era voltado ao mercado de locação. Bela Vista (índice 100), Rio Branco (57) e Farroupilha (67) permanecem entre os de maior valorização em aluguéis de 2026, enquanto Rubem Berta (índice 5) registra o segundo maior desconto (−0,818; $p < 0,001$) e Belém Velho — ausente da amostra de 1992 — lidera os descontos (−0,854; $p < 0,001$). A amplitude entre Jardim Europa (+0,548) e Belém Velho (−0,854) equivale a uma diferença de aproximadamente 306% no aluguel de dois imóveis com atributos idênticos.

Os diagnósticos do modelo de aluguel seguiram os mesmos procedimentos aplicados ao modelo de vendas. Os VIFs das variáveis contínuas situaram-se abaixo de 2,8, sem indicação de multicolinearidade. O teste de Breusch-Pagan resultou em BP = 520,37 ($p < 0,001$), e erros robustos HC1 foram adotados.

Figura 5 - Resíduos vs Valores ajustados do modelo de aluguel

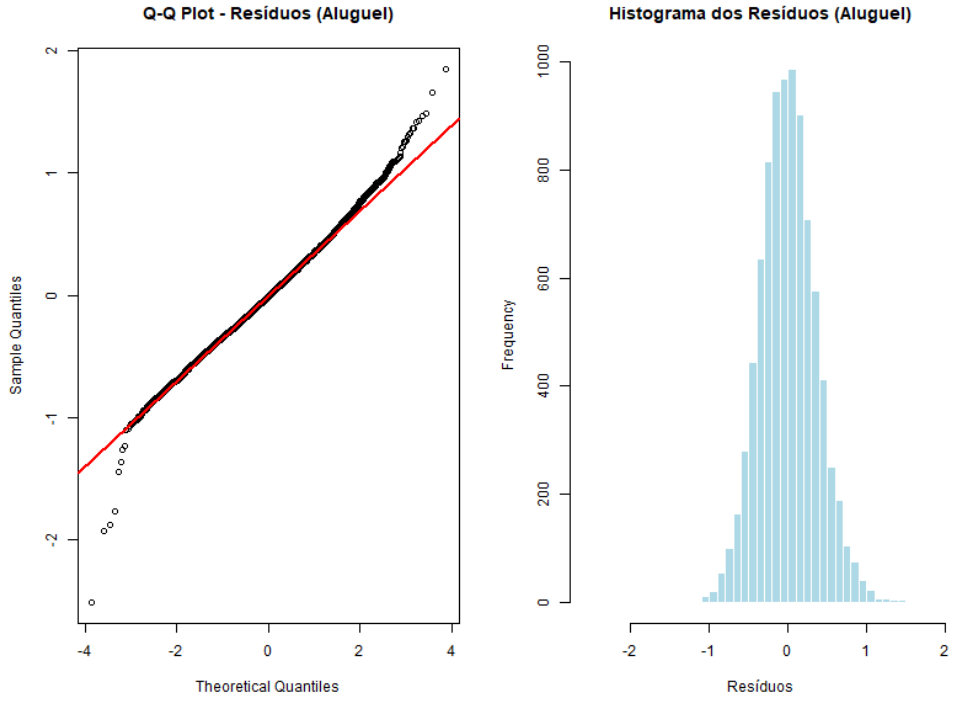


Fonte: Elaboração Própria (2026)

A Figura 5 apresenta os resíduos contra os valores ajustados do modelo de aluguel. Diferentemente do padrão observado para vendas, nota-se um agrupamento em faixas horizontais — linhas de pontos alinhadas em alturas específicas. Esse padrão é característico de dados com arredondamento nos preços: aluguéis são frequentemente anunciados em

valores redondos (R$ 1.500, R$ 2.000, R$ 2.500), o que gera concentração de resíduos em intervalos discretos quando o valor ajustado pelo modelo não coincide com esses arredondamentos. Embora visualmente distinto, esse fenômeno não invalida a inferência sob erros robustos HC1.

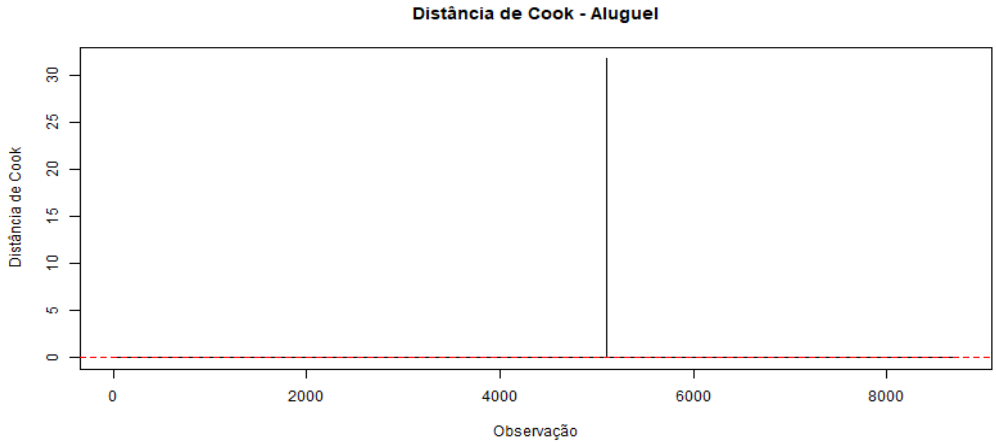
Figura 6 - Q-Q Plot e Histograma dos Resíduos do modelo de aluguel



Fonte: Elaboração Própria (2026)

Figura 6 exibe o Q-Q plot e o histograma dos resíduos do modelo de aluguel. No painel esquerdo, os pontos acompanham a reta teórica na região central, com desvios mais pronunciados na cauda superior — resíduos positivos acima do esperado pela distribuição normal —, indicando a presença de imóveis com aluguéis acima do previsto pelo modelo. No painel direito, o histograma revela distribuição com leve assimetria à direita, consistente com o observado no Q-Q.

Figura 7 - Distância de Cook do modelo de aluguel



Fonte: Elaboração Própria (2026)

A Figura 7 exibe a Distância de Cook para o modelo de aluguel. O padrão é análogo ao de vendas: a imensa maioria das observações situa-se abaixo do limiar, com poucos picos extremos. O mais expressivo, próximo à observação 5.000, apresenta D > 30 — valor excepcionalmente elevado, provavelmente correspondente a um anúncio com combinação de atributos radicalmente atípica (por exemplo, um aluguel muito alto em bairro periférico, ou um imóvel de grande porte com poucas amenidades). A remoção dessas observações influentes eliminou aproximadamente 4,6% da amostra (de 8.737 para 8.338 observações) e esteve associada a uma elevação do R^2 ajustado de 0,745 para 0,781.

4.3 SÍNTESE COMPARATIVA

A comparação entre os dois modelos revela padrões consistentes e divergências economicamente interpretáveis.

Tabela – 7: Comparação dos resultados

Variável	Coef (venda)	Sig.	Coef (aluguel)	Sig.
Log(área)	0,577	***	0,406	***
Dormitórios	-0,000	NS	0,036	***
Suítes	0,237	***	0,213	***
Churrasqueira	0,141	***	0,144	***
Elevador	0,061	***	0,133	***
Piscina	0,138	***	0,086	***

Fonte: Elaboração Própria (2026)

A divergência mais expressiva entre os dois modelos reside na elasticidade-área: 0,577 para vendas contra 0,406 para aluguéis. Em termos concretos, as evidências demonstram que uma duplicação na área de um imóvel pode gerar acréscimo de 49% no preço de venda ($2^{0,577} - 1$), mas de apenas 32% no aluguel mensal ($20^{0,406-1}$). Essa assimetria é consistente com duas lógicas econômicas distintas. Para o comprador, o metro quadrado adicional pode demonstrar simultaneamente consumo (espaço de moradia) e investimento (ativo patrimonial que acompanha a valorização do mercado); para o locatário, o foco pode recair sobre o consumo, cujo benefício precisa justificar o custo mensal recorrente.

A variável dormitórios é não significativa no modelo de vendas, mas positiva e significativa no de aluguéis ($\beta = 0,036$; $p < 0,01$). No mercado de compra, a presença de log(área) absorve o efeito de dimensionamento: imóveis com mais quartos são maiores, e o tamanho já está captado. No mercado de locação, a configuração funcional pode operar como critério primário de busca. González (1993, p. 89) reportava colinearidade entre dormitórios e área em seu modelo, mas utilizava a variável QUARTOS em vez de ÁREA em parte de suas especificações, obtendo coeficientes significativos.

Os coeficientes de suítes e piscina são notavelmente estáveis entre os dois mercados: suítes associam-se a prêmio de 26,7% em vendas e 23,8% em aluguéis; piscina, 15,1% e 15,4%. Essa estabilidade é teoricamente informativa. Na linguagem de Rosen (1974), significa que o preço-sombra desses atributos é aproximadamente o mesmo nas funções de lance de compradores e locatários — ambos os grupos atribuem valor similar a essas características. A interpretação mais parcimoniosa é que suítes e piscina operam como sinalizadores de padrão construtivo cujo valor informacional independe do horizonte da transação. Esse resultado valida, com dados de 2026, a intuição de González (1993, p. 48) sobre a importância de captar qualidade construtiva: as variáveis que melhor cumprem esse papel são precisamente aquelas cujo preço-sombra se mantém constante entre mercados.

As maiores assimetrias aparecem em churrasqueira e elevador. A churrasqueira está associada a prêmio de 14,2% no aluguel, mas apenas 6,3% na venda — um diferencial de 125%. O elevador segue caminho inverso: 14,8% na venda, 9,0% no aluguel. Essas divergências podem ser interpretáveis à luz dos horizontes de decisão. O locatário, cujo contrato é temporário e possui menos possibilidades de alteração no imóvel, valoriza atributos de fruição imediata. A churrasqueira será utilizada já no próximo fim de semana. O comprador, que projeta um uso prolongado, pode privilegiar atributos estruturais associados à acessibilidade e à valorização futura.

A hierarquia espacial de bairros é qualitativamente consistente entre vendas e aluguéis: Jardim Europa, Três Figueiras, Bela Vista e Moinhos de Vento lideram os prêmios em ambos os mercados. Essa consistência é teoricamente esperada: na formulação de Rosen (1974), o preço-sombra da localização reflete o equilíbrio entre a disposição a pagar dos consumidores e o custo de oportunidade dos produtores, e ambos os mecanismos operam tanto na venda quanto na locação. A comparação com González (1993, Tabela 5.12), que classificava Bela Vista no topo e Rubem Berta na base de seu índice locacional, demonstra que essa hierarquia é não apenas transversal aos mercados, mas também estável ao longo de três décadas de transformações urbanas em Porto Alegre.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa propôs-se a revisitar as hipóteses de González (1993) sobre a formação de preços no mercado imobiliário de Porto Alegre, substituindo a coleta presencial de 504 imóveis por bases digitais — 68.051 anúncios de venda e 8.338 de locação extraídos de portais online. O principal resultado é a estabilidade da estrutura de preços hedônicos da cidade: três décadas depois, e a partir de dados de origem distinta, área e localização permanecem como os atributos de maior peso na formação do valor, e a ordenação dos bairros se mantém. A esse resultado somam-se duas contribuições — o uso de amenidades de anúncio como proxies de qualidade e a comparação entre os mercados de venda e locação.

A replicação reproduziu os padrões identificados por González. Área e localização figuram como os atributos mais associados ao preço dos imóveis, com retornos decrescentes da metragem, e o ranking de prêmios locacionais por bairro manteve-se ao longo de três décadas: entre os bairros já existentes em 1993, Bela Vista, Rio Branco e Moinhos de Vento permanecem no topo e Rubem Berta e Sarandi na base, ordenação próxima à registrada pelo autor. Bairros mais recentes, como Jardim Europa — hoje entre os de maior prêmio (Seção 4) —, não têm correspondente naquele período e, por isso, ficam fora dessa comparação intertemporal. A transição de dados analógicos para digitais não reduziu o poder explicativo dos modelos, que alcançaram R^2 de 0,836 para vendas e 0,783 para aluguéis, com validade aferida por um conjunto de diagnósticos econométricos: ausência de multicolinearidade (VIFs abaixo de 2,6), correção da heterocedasticidade por erros-padrão robustos HC1 e depuração de observações influentes pela Distância de Cook, que removeu cerca de 4% da amostra de vendas e 4,6% da de aluguéis.

A segunda contribuição diz respeito às amenidades observáveis em anúncios. Suítes, piscina, churrasqueira e elevador apresentaram coeficientes positivos e significativos nos dois mercados, e a presença de suítes registrou o maior prêmio entre elas — 26,7% na venda e 23,8% no aluguel. Esses atributos cumpriram papel equivalente ao dos índices de padrão construtivo e estado de conservação que González (1993, p. 48) registrava em vistoria, o que indica que a precificação hedônica pode prescindir da inspeção presencial quando se dispõe de descritores de qualidade nos próprios anúncios.

A comparação entre venda e locação revelou assimetrias entre os dois mercados. No segmento de venda, a metragem entra com maior peso (elasticidade de 0,577 contra 0,406) e o

elevador recebe prêmio maior (14,8%, ante 9,0% na locação), atributos ligados ao uso prolongado e à valorização futura. Nos anúncios de locação, a churrasqueira (14,2%, contra 6,3% na venda) e o número de dormitórios recebem prêmios maiores, atributos de uso imediato, associados ao horizonte mais curto do contrato. Suítes e piscina mantêm prêmios semelhantes nos dois segmentos, o que indica que seu papel de sinalização de qualidade independe do tipo de transação.

Cabe reconhecer as limitações do estudo. A base é composta por preços de oferta, e não de transações efetivas, o que pode introduzir um viés de sobrevalorização, já que os valores anunciados tendem a superar os de fechamento. Além disso, a natureza de corte transversal não permite captar dinâmicas temporais. Por fim, cabe destacar a principal limitação da estratégia empírica: os resultados expressam correlações parciais, não relações de causalidade. Nesse sentido, trabalhos futuros usando dados em painel com efeitos fixos de unidade e de tempo são um caminho natural da pesquisa empírica.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974. DOI: https://doi.org/10.1109/TAC.1974.1100705. Acesso em: 10 mar. 2026.

ARRAES, R. A.; SOUSA FILHO, E. Externalidades e formação de preços no mercado imobiliário urbano brasileiro: um estudo de caso. **Economia Aplicada**, Ribeirão Preto, v. 12, n. 2, p. 289-319, 2008. DOI: https://doi.org/10.1590/S1413-80502008000200006. Acesso em 10 out. 2025.

BALCHIN, P. N.; KIEVE, J. L. **Urban land economics**. 3. ed. London: Macmillan, 1986. Disponível em: https://nla.gov.au/nla.cat-vn2356111. Acesso em: 10 abr. 2026.

BELSLEY, D. A.; KUH, E.; WELSCH, R. E. **Regression diagnostics: identifying influential data and sources of collinearity**. New York: John Wiley & Sons, 1980. DOI: https://doi.org/10.1002/0471725153. Acesso em: 11 abr. 2026.

BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**, v. 47, n. 5, p. 1287-1294, 1979. DOI: https://doi.org/10.2307/1911963. Acesso em: 12 mar. 2026.

BROWN, J. N.; ROSEN, H. S. On the estimation of structural hedonic price models. Econometrica, v. 50, n. 3, p. 765-768, 1982. DOI: https://doi.org/10.2307/1912614. Acesso em: 12 jun. 2025.

CAMPOS, R. B. A. O mercado imobiliário residencial no município de São Paulo: uma abordagem de preços hedônicos espacial. **Nova Economia**, v. 27, n. 1, p. 303-337, 2017. DOI: https://doi.org/10.1590/0103-6351/2441. Acesso em: 01 jun. 2026.

COOK, R. D. Detection of influential observations in linear regression. **Technometrics**, v. 19, n. 1, p. 15-18, 1977. DOI: https://doi.org/10.1080/00401706.1977.10489493. Acesso em: 01 mar. 2026.

FIPE; ZAP. **Índice FipeZAP de Preços de Imóveis Anunciados**. São Paulo: Fipe, 2025; 2026. Disponível em: https://www.fipe.org.br/pt-br/indices/fipezap/. Acesso em: 7 jun. 2026.

FREEMAN, A. M. The benefits of environmental improvement: theory and practice. Baltimore: **Johns Hopkins University Press**, 1979. Disponível em: https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=2897&context=nrj. Acesso em: 06 jun. 2026.

GONZÁLEZ, M. A. S. **A formação do valor dos aluguéis de apartamentos residenciais na cidade de Porto Alegre.** 1993. Dissertação (Mestrado em Engenharia) – Curso de Pós-Graduação em Engenharia Civil, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993. Disponível em: http://lume.ufrgs.br/handle/10183/189661. Acesso em: 10 jun. 2025.

GONZÁLEZ, M. A. S.; FORMOSO, C. T. **Estimativa de modelos de preços hedônicos para locação residencial em Porto Alegre**. Produção, v. 5, n. 1, p. 65-77, 1995. DOI: https://doi.org/10.1590/S0103-65131995000100004. Acesso em: 10 jun. 2025.

HALVORSEN, R.; PALMQUIST, R. The interpretation of dummy variables in semilogarithmic equations. **American Economic Review**, v. 70, n. 3, p. 474-475, 1980. Disponível em: https://scispace.com/pdf/the-interpretation-of-dummy-variables-in-semilogarithmic-54qeaggyyn.pdf. Acesso em: 12 mar. 2026.

HARRISON, D.; RUBINFELD, D. L. Hedonic housing prices and the demand for clean air. **Journal of Environmental Economics and Management**, v. 5, n. 1, p. 81-102, 1978. DOI: https://doi.org/10.1016/0095-0696(78)90006-2. Acesso em: 20 mar. 2025.

HERMANN, B. M.; HADDAD, E. A. Mercado imobiliário e amenidades urbanas: a view through the window. **Estudos Econômicos** (São Paulo), v. 35, n. 2, p. 237-269, 2005. Disponível em: https://www.revistas.usp.br/ee/article/view/35841. Acesso em: 15 mar. 2025.

IBGE. **Pesquisa de Orçamentos Familiares 2017-2018**: primeiros resultados. Rio de Janeiro: IBGE, 2019. Disponível em: https://basedosdados.org/dataset/a1b6d2b6-4aa6-47e7-a517-8a21b28b7254?table=7b880731-ffa2-4bde-a290-ae058b3acf51. Acesso em: 5 jun. 2026.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2022**: primeiros resultados. Rio de Janeiro: IBGE, 2022. Disponível em: https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102011. Acesso em: 15 jan. 2026.

JULIO, R. O. B. **Modelo de preços hedônicos espacial:** análise dos mercados implícitos para os imóveis residenciais de Porto Alegre. 2021. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: http://hdl.handle.net/10183/234785. Acesso em: 05 jun. 2026.

LANCASTER, K. J. A new approach to consumer theory. **Journal of Political Economy**, v. 74, n. 2, p. 132-157, 1966. DOI: https://doi.org/10.1086/259131. Acesso em: 10 jun. 2025.

MACKINNON, J. G.; WHITE, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. **Journal of Econometrics**, v. 29, n. 3, p. 305-325, 1985. DOI: https://doi.org/10.1016/0304-4076(85)90158-7. Acesso em: 29 mar. 2026.

PORTO ALEGRE. **Lei Complementar nº 434**, de 1º de dezembro de 1999. Institui o Plano Diretor de Desenvolvimento Urbano Ambiental (PDDUA). Porto Alegre: Câmara Municipal, 1999. Disponível em: https://lproweb.procempa.com.br/pmpa/prefpoa/edificapoa/usu_doc/pddua180113.pdf. Acesso em: 01 jun. 2026.

PORTO ALEGRE. **Lei Complementar nº 646**, de 22 de julho de 2010. Revisão do PDDUA. Porto Alegre: Câmara Municipal, 2010. Disponível em: https://www2.portoalegre.rs.gov.br/cgi-bin/nph-brs?s1=000031194.DOCN.&l=20&u=/netahtml/sirel/simples.html&p=1&r=1&f=G&d=atos&SECT1=TEXT. Acesso em: 05 jun. 2026.

PORTO ALEGRE. **Plano Diretor Urbano Sustentável** (PDUS). PLCE 019/25, aprovado em 23 de abril de 2026. Porto Alegre: Câmara Municipal, 2026. Disponível em:

https://www.camarapoa.rs.gov.br/draco/processos/141634/PLCE_20-25.pdf. Acesso em: 06 jun. 2026.

ROSEN, S. Hedonic prices and implicit markets: product differentiation in pure competition. **Journal of Political Economy**, v. 82, n. 1, p. 34-55, 1974. DOI: https://doi.org/10.1086/260169. Acesso em: 07 mar. 2025.

SEABRA, D. M. S.; SILVEIRA NETO, R. M.; MENEZES, T. A. Amenidades urbanas e valor das residências: uma análise empírica para a cidade do Recife. **Economia Aplicada**, v. 20, n. 1, p. 143-169, 2016. DOI: https://doi.org/10.11606/1413-8050/ea147474. Acesso em: 01 jun. 2026.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591-611, 1965. DOI: https://doi.org/10.1093/biomet/52.3-4.591. Acesso em: 06 mar. 2025.

SHEPPARD, S. Hedonic analysis of housing markets. In: CHESHIRE, P.; MILLS, E. S. (ed.). **Handbook of Regional and Urban Economics**. Amsterdam: Elsevier, v. 3, p. 1595-1635, 1999. DOI: https://doi.org/10.1016/S1574-0080(99)80010-8. Acesso em: 19 maio 2025.

VECCHIA, I. D. **Modelagem hedônica e econometria espacial:** predição de preços de imóveis comerciais em Porto Alegre-RS. 2025. Trabalho de Conclusão de Curso (Graduação em Administração) – Escola de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2025. Disponível em: http://hdl.handle.net/10183/294757. Acesso em: 05 jun. 2026.

WHITE, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. **Econometrica**, v. 48, n. 4, p. 817-838, 1980. DOI: https://doi.org/10.2307/1912934. Acesso em: 30 mar. 2026

WOOLDRIDGE, J. M. **Introdução à econometria:** uma abordagem moderna. Tradução de Heloisa Pinna Bernardo. 4. ed. São Paulo: Cengage Learning, 2023. E-book.

APÊNDICE A – FOLHA DE APROVAÇÃO

LORENZZO SOARES SANTOS

PREÇOS DOS IMÓVEIS EM PORTO ALEGRE EM 2026: UMA ANÁLISE VIA MODELO DE PREÇOS HEDÔNICOS

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Aprovada em: Porto Alegre, _____de _____ de 2026.

BANCA EXAMINADORA:

Prof. Dr. Sabino da Silva Porto Júnior – Orientador
UFRGS

Prof. Dr. Fabian Scholze Domingues
UFRGS

Prof. Dr. Eduardo André Tillmann
FURG

APÊNDICE B – COEFICIENTES DE LOCALIZAÇÃO DO MODELO DE VENDAS

Variável	Coef	Std.err	t_valor	Sig
Anchieta	-0,2889	0,2403	-1,2026	NS
Arquipélago	0,0720	0,2686	0,2682	NS
Auxiliadora	0,4030	0,1702	2,3682	*
Azenha	0,0510	0,1701	0,3001	NS
Bela Vista	0,4831	0,1701	2,8391	**
Belém Velho	-0,2456	0,1739	-1,4125	NS
Boa Vista	0,3508	0,1701	2,0617	*
Bom Fim	0,3031	0,1702	1,7805	.
Bom Jesus	-0,0964	0,1706	-0,5649	NS
Camaquã	-0,1802	0,1702	-1,0588	NS
Campo Novo	-0,3303	0,2031	-1,6263	NS
Cascata	-0,2694	0,1917	-1,4053	NS
Cavalhada	-0,1755	0,1701	-1,0317	NS
Centro Histórico	0,0151	0,1700	0,0887	NS
Chácara das Pedras	0,2690	0,1702	1,5801	NS
Cidade Baixa	0,2351	0,1700	1,3829	NS
Cor Aparício Borges	-0,3088	0,1811	-1,7049	.
Costa e Silva	-0,2407	0,1723	-1,3967	NS
Cristal	-0,0611	0,1700	-0,3594	NS
Cristo Redentor	0,0492	0,1701	0,2890	NS
Farrapos	-0,2698	0,1715	-1,5735	NS
Farroupilha	0,3102	0,1733	1,7906	.
Floresta	-0,0263	0,1702	-0,1544	NS
Glória	-0,1651	0,1704	-0,9685	NS
Higienópolis	0,2451	0,1701	1,4408	NS
Humaitá	-0,2899	0,1704	-1,7015	.
Independência	0,1907	0,1702	1,1206	NS
Ipanema	-0,1026	0,1708	-0,6003	NS
Jardim Botânico	0,1634	0,1701	0,9608	NS
Jardim Carvalho	-0,0833	0,1702	-0,4897	NS
Jardim do Salso	0,1663	0,1703	0,9769	NS
Jardim Europa	0,6460	0,1708	3,7836	***
Jardim Floresta	-0,1130	0,1722	-0,6561	NS
Jardim Isabel	-0,0724	0,1706	-0,4242	NS
Jardim Itu	-0,0768	0,1701	-0,4514	NS
Jardim Leopoldina	-0,3566	0,1705	-2,0913	*
Jardim Lindóia	0,1154	0,1702	0,6782	NS
Jardim Sabará	-0,0844	0,1703	-0,4957	NS
Jardim São Pedro	-0,0488	0,1710	-0,2854	NS
Lomba do Pinheiro	-0,2479	0,1705	-1,4538	NS
Mário Quintana	-0,1266	0,1704	-0,7430	NS
Medianeira	-0,1296	0,1702	-0,7616	NS
Menino Deus	0,2311	0,1700	1,3597	NS
Moinhos de Vento	0,4715	0,1701	2,7716	**
Montserrat	0,4283	0,1701	2,5178	*
Morro Santana	-0,2117	0,1702	-1,2441	NS
Navegantes	-0,1400	0,1713	-0,8168	NS
Nonoai	-0,2116	0,1701	-1,2440	NS
Parque Santa Fé	-0,3436	0,1706	-2,0136	*
Partenon	-0,0451	0,1700	-0,2651	NS
Passo da Areia	0,2096	0,1700	1,2330	NS
Passo das Pedras	-0,2717	0,1711	-1,5880	NS
Pedra Redonda	0,1379	0,1722	0,8005	NS
Petrópolis	0,3384	0,1700	1,9911	*

Praia de Belas	0,3354	0,1708	1,9635	*
Rio Branco	0,3922	0,1700	2,3065	*
Rubem Berta	-0,4229	0,1710	-2,4727	*
Santa Cecília	0,0942	0,1704	0,5527	NS
Santa Maria Goretti	-0,0666	0,1710	-0,3897	NS
Santa Rosa de Lima	-0,4414	0,1718	-2,5697	*
Santa Tereza	-0,0804	0,1702	-0,4725	NS
Santana	0,1111	0,1701	0,6534	NS
Santo Antônio	-0,1155	0,1702	-0,6788	NS
São Geraldo	-0,0622	0,1703	-0,3651	NS
São João	0,0902	0,1702	0,5300	NS
São Sebastião	0,0365	0,1702	0,2142	NS
Sarandi	-0,1716	0,1700	-1,0090	NS
Sétimo Céu	-0,0450	0,1732	-0,2596	NS
Teresópolis	-0,0763	0,1701	-0,4483	NS
Três Figueiras	0,5514	0,1702	3,2391	**
Tristeza	0,1121	0,1700	0,6596	NS
Vila Assunção	0,1734	0,1706	1,0168	NS
Vila Conceição	0,1186	0,1743	0,6806	NS
Vila Ipiranga	0,0386	0,1701	0,2268	NS
Vila Jardim	0,1432	0,1706	0,8394	NS
Vila João Pessoa	-0,2055	0,1725	-1,1914	NS
Vila Nova	-0,2526	0,1701	-1,4852	NS
Vila São José	-0,2524	0,1712	-1,4741	NS

Fonte: Elaboração Própria (2026)

APÊNDICE C – COEFICIENTES DE LOCALIZAÇÃO DO MODELO DE ALUGUEL

Variável	Coef	Std.err	t_valor	Sig
Anchieta	-0,4701	0,2733	-1,7198	.
Auxiliadora	0,2997	0,1484	2,0195	*
Azenha	-0,1932	0,1494	-1,2935	NS
Bela Vista	0,3051	0,1483	2,0577	*
Belém Velho	-0,8543	0,1917	-4,4563	***
Boa Vista	0,2245	0,1489	1,5073	NS
Bom Fim	-0,0490	0,1499	-0,3269	NS
Bom Jesus	-0,2429	0,1539	-1,5779	NS
Camaquã	-0,4055	0,1532	-2,6474	**
Campo Novo	-0,6116	0,3575	-1,7107	.
Cascata	-0,2098	0,2733	-0,7676	NS
Cavalhada	-0,4142	0,1500	-2,7623	**
Centro Histórico	-0,2668	0,1475	-1,8094	.
Chácara das Pedras	0,0537	0,1504	0,3574	NS
Cidade Baixa	-0,1654	0,1481	-1,1166	NS
Cor. Aparício Borges	-0,5622	0,3573	-1,5732	NS
Costa e Silva	-0,5224	0,1558	-3,3532	***
Cristal	-0,2995	0,1493	-2,0063	*
Cristo Redentor	-0,1743	0,1488	-1,1708	NS
Farrapos	-0,4300	0,1540	-2,7918	**
Farroupilha	-0,0638	0,1647	-0,3873	NS
Floresta	-0,3353	0,1501	-2,2334	*
Glória	-0,5939	0,1551	-3,8292	***
Higienópolis	0,0034	0,1492	0,0228	NS
Humaitá	-0,5449	0,1515	-3,5970	***
Independência	-0,1396	0,1499	-0,9314	NS
Ipanema	-0,1911	0,1825	-1,0474	NS
Jardim Botânico	-0,0583	0,1491	-0,3910	NS
Jardim Carvalho	-0,2561	0,1494	-1,7141	.
Jardim do Salso	-0,0600	0,1501	-0,3993	NS
Jardim Europa	0,5482	0,1524	3,5970	***
Jardim Floresta	-0,3189	0,1722	-1,8520	.
Jardim Isabel	-0,2816	0,1673	-1,6831	.
Jardim Itu	-0,2206	0,1493	-1,4772	NS
Jardim Leopoldina	-0,6463	0,1592	-4,0586	***
Jardim Lindóia	-0,0985	0,1509	-0,6526	NS
Jardim Sabará	-0,4103	0,1528	-2,6856	**
Jardim São Pedro	-0,3229	0,1591	-2,0293	*
Lomba do Pinheiro	-0,5220	0,1561	-3,3430	***
Mário Quintana	0,0476	0,1505	0,3163	NS
Medianeira	-0,4367	0,1511	-2,8898	**
Menino Deus	-0,1021	0,1478	-0,6913	NS
Moinhos de Vento	0,2843	0,1479	1,9216	.
Montserrat	0,3156	0,1484	2,1265	*
Morro Santana	-0,4810	0,1510	-3,1844	**
Navegantes	-0,4385	0,1545	-2,8386	**
Nonoai	-0,4001	0,1502	-2,6636	**
Parque Santa Fé	-0,6610	0,1619	-4,0842	***
Partenon	-0,1715	0,1480	-1,1587	NS
Passo da Areia	-0,0259	0,1481	-0,1747	NS
Passo das Pedras	-0,2743	0,1792	-1,5306	NS
Pedra Redonda	0,3392	0,1574	2,1544	*
Petrópolis	0,1201	0,1473	0,8153	NS
Praia de Belas	-0,0003	0,1535	-0,0018	NS
Rio Branco	0,2377	0,1480	1,6064	NS
Rubem Berta	-0,8182	0,1585	-5,1606	***

Santa Cecília	-0,1458	0,1532	-0,9516	NS
Santa Maria Goretti	-0,3295	0,1542	-2,1373	*
Santa Rosa de Lima	-0,6022	0,1916	-3,1425	**
Santa Tereza	-0,4165	0,1515	-2,7484	**
Santana	-0,1217	0,1483	-0,8204	NS
Santo Antônio	-0,3438	0,1506	-2,2836	*
São Geraldo	-0,3580	0,1503	-2,3824	*
São João	-0,1833	0,1503	-1,2202	NS
São Sebastião	-0,2153	0,1525	-1,4119	NS
Sarandi	-0,3811	0,1486	-2,5646	*
Teresópolis	-0,3015	0,1498	-2,0131	*
Três Figueiras	0,3820	0,1487	2,5684	*
Tristeza	-0,1550	0,1488	-1,0415	NS
Vila Assunção	0,1372	0,1538	0,8919	NS
Vila Conceição	-0,1308	0,2731	-0,4789	NS
Vila Ipiranga	-0,1844	0,1508	-1,2234	NS
Vila Jardim	0,0955	0,1520	0,6284	NS
Vila João Pessoa	-0,4068	0,1629	-2,4966	*
Vila Nova	-0,5707	0,1507	-3,7860	***
Vila São José	-0,4890	0,1917	-2,5506	*

Fonte: Elaboração Própria (2026)

APÊNDICE D – Script em R utilizado para a modelagem dos dados

```
# TCC: Preços dos Imóveis em Porto Alegre em 2026
#       Uma Análise via Modelo de Preços Hedônicos
# Autor: Lorenzzo Soares Santos
# Orientador: Dr. Sabino da Silva Porto Junior
# UFRGS - Faculdade de Ciências Econômicas
#
================================================================================
==
# Script completo: limpeza, estimação, diagnósticos e exportação
#
================================================================================
 ================================================================================
# 0. PACOTES NECESSÁRIOS
# ================================================================================
# Instalar apenas uma vez:
#  install.packages(c("car", "lmtest", "sandwich", "stargazer", "writexl",
"ggplot2"))

library(car)        # vif() - teste de multicolinearidade
library(lmtest)               #   bptest()   -   teste   de   Breusch-Pagan
(heterocedasticidade)
library(sandwich)   # vcovHC() - erros-padrão robustos (HC1)
library(stargazer)  # tabelas formatadas para o TCC
library(writexl)    # exportar para Excel
library(ggplot2)    # gráficos
# ================================================================================
# 1. CARREGAR AS BASES DE DADOS
# ================================================================================
cat("=== ETAPA 1: Carregando dados ===\n")
setwd("C:/Users/loren/OneDrive/Documentos/Base de dados tcc")
# Carregar os dados diretamente (reprodutível)
dados_venda_brutos  <- read.csv("porto_alegre_venda_zap.csv.csv")
dados_aluguel_brutos <- read.csv("porto_alegre_aluguel_zap.csv.csv")

cat("Vendas brutas:", nrow(dados_venda_brutos), "obs\n")
cat("Aluguel bruto:", nrow(dados_aluguel_brutos), "obs\n")
# ================================================================================
# 2. FUNÇÕES AUXILIARES DE LIMPEZA
# ================================================================================
# Função para limpar campos entre colchetes: "[3]" -> 3, "[]" -> 0
limpar_colchetes <- function(x) {
  num <- as.numeric(gsub("\\[|\\]|'", "", as.character(x)))
  num[is.na(num)] <- 0
  return(num)
}
# Função para extrair preço do campo pricingInfos (se necessário)
# Se preco_venda e preco_aluguel já existem, pular

# ================================================================================
# 3. LIMPEZA DA BASE DE VENDAS
# ================================================================================
cat("\n=== ETAPA 3: Limpeza - Base de Vendas ===\n")

dados_v <- dados_venda_brutos

# 3a. Limpar variáveis numéricas que vêm como texto com colchetes
dados_v$quartos   <- limpar_colchetes(dados_v$bedrooms)
dados_v$banheiros <- limpar_colchetes(dados_v$bathrooms)
```

```
dados_v$suites    <- limpar_colchetes(dados_v$suites)
dados_v$vagas     <- limpar_colchetes(dados_v$parkingSpaces)

# 3b. Filtros básicos de qualidade
dados_v <- subset(dados_v,
                  area > 10 &              # mínimo 10m² (exclui erros)
                    area < 1500 &            # máximo razoável
                    preco_venda > 50000 &    # mínimo R$50k
                    preco_venda < 20000000 & # máximo R$20M
                    quartos >= 0 &
                    vagas >= 0 &
                    tipologia  !=  "RESIDENTIAL_ALLOTMENT_LAND"  &   # exclui
terrenos
                    name != "" & !is.na(name))  # bairro identificado

cat("Após filtros básicos:", nrow(dados_v), "obs\n")

# 3c. Remover duplicatas (mesmo local, área, quartos e preço)
dados_v <- dados_v[!duplicated(dados_v[c("latitude", "longitude",
                                         "area",                 "quartos",
"preco_venda")]), ]
cat("Após remoção de duplicatas:", nrow(dados_v), "obs\n")

# 3d. Criar variáveis dummy de amenidades
dados_v$tem_piscina            <-  ifelse(grepl("POOL",  dados_v$amenities,
ignore.case = TRUE), 1, 0)
dados_v$tem_churrasqueira  <-  ifelse(grepl("BARBECUE",  dados_v$amenities,
ignore.case = TRUE), 1, 0)
dados_v$tem_elevador        <-  ifelse(grepl("ELEVATOR",  dados_v$amenities,
ignore.case = TRUE), 1, 0)

# 3e. Criar preço por m²
dados_v$preco_m2 <- dados_v$preco_venda / dados_v$area

# 3f. Remoção de outliers pelo método IQR (percentis 1% e 99%)
#     Método transparente e reprodutível
q_preco <- quantile(dados_v$preco_venda, c(0.01, 0.99))
q_area  <- quantile(dados_v$area, c(0.01, 0.99))

dados_v <- subset(dados_v,
                  preco_venda >= q_preco[1] & preco_venda <= q_preco[2] &
                    area >= q_area[1] & area <= q_area[2])

cat("Após remoção de outliers (percentil 1%-99%):", nrow(dados_v), "obs\n")

# ============================================================================
# 4. LIMPEZA DA BASE DE ALUGUEL
# ============================================================================
cat("\n=== ETAPA 4: Limpeza - Base de Aluguel ===\n")

dados_a <- dados_aluguel_brutos

dados_a$quartos   <- limpar_colchetes(dados_a$bedrooms)
dados_a$banheiros <- limpar_colchetes(dados_a$bathrooms)
dados_a$suites    <- limpar_colchetes(dados_a$suites)
dados_a$vagas     <- limpar_colchetes(dados_a$parkingSpaces)

dados_a <- subset(dados_a,
                  area > 10 & area < 1500 &
                    preco_aluguel > 200 &       # mínimo R$200/mês
                    preco_aluguel < 50000 &     # máximo R$50k/mês
```

```
                    quartos >= 0 &
                    vagas >= 0 &
                    tipologia != "RESIDENTIAL_ALLOTMENT_LAND" &
                    name != "" & !is.na(name))

dados_a <- dados_a[!duplicated(dados_a[c("latitude", "longitude",
                                         "area",                "quartos",
"preco_aluguel")]), ]

dados_a$tem_piscina            <-  ifelse(grepl("POOL",  dados_a$amenities,
ignore.case = TRUE), 1, 0)
dados_a$tem_churrasqueira  <-  ifelse(grepl("BARBECUE",  dados_a$amenities,
ignore.case = TRUE), 1, 0)
dados_a$tem_elevador        <-  ifelse(grepl("ELEVATOR",  dados_a$amenities,
ignore.case = TRUE), 1, 0)

q_aluguel <- quantile(dados_a$preco_aluguel, c(0.01, 0.99))
q_area_a  <- quantile(dados_a$area, c(0.01, 0.99))
dados_a <- subset(dados_a,
                  preco_aluguel   >=   q_aluguel[1]   &   preco_aluguel   <=
q_aluguel[2] &
                    area >= q_area_a[1] & area <= q_area_a[2])

cat("Aluguel final:", nrow(dados_a), "obs\n")

# ============================================================================
# 5. ESTATÍSTICAS DESCRITIVAS
# ============================================================================
cat("\n=== ETAPA 5: Estatísticas Descritivas ===\n")

vars_descritivas_v   <-   dados_v[,   c("preco_venda",   "area",   "quartos",
"banheiros",
                                  "suites", "vagas", "tem_piscina",
                                  "tem_churrasqueira", "tem_elevador")]
cat("\n--- Vendas ---\n")
print(summary(vars_descritivas_v))

vars_descritivas_a  <-  dados_a[,  c("preco_aluguel",  "area",  "quartos",
"banheiros",
                                  "suites", "vagas", "tem_piscina",
                                  "tem_churrasqueira", "tem_elevador")]
cat("\n--- Aluguel ---\n")
print(summary(vars_descritivas_a))

# ============================================================================
# 6. ESTIMAÇÃO DO MODELO HEDÔNICO - VENDAS
# ============================================================================
cat("\n=== ETAPA 6: Modelo Hedônico de Vendas ===\n")

# ----------------------------------------------------------------------------
-
# MODELO 1 (Básico): log(P) = f(area, quartos, vagas, bairro, tipologia)
# Justificativa: forma funcional log-linear, padrão na literatura
# (Rosen, 1974; González, 1993; Hermann & Haddad, 2005)
# ----------------------------------------------------------------------------
-
modelo_v1 <- lm(log(preco_venda) ~ area + quartos + vagas +
                  as.factor(tipologia) + as.factor(name),
                data = dados_v)

cat("\n--- Modelo Vendas 1 (Básico) ---\n")
```

```
cat("R² ajustado:", summary(modelo_v1)$adj.r.squared, "\n")
cat("N° observações:", nobs(modelo_v1), "\n")

# ------------------------------------------------------------------------
-
# MODELO 2 (Avançado): adiciona amenidades
# ------------------------------------------------------------------------
-
modelo_v2 <- lm(log(preco_venda) ~ area + quartos + vagas + suites +
                  tem_piscina + tem_churrasqueira + tem_elevador +
                  as.factor(tipologia) + as.factor(name),
                data = dados_v)

cat("\n--- Modelo Vendas 2 (Avançado) ---\n")
cat("R² ajustado:", summary(modelo_v2)$adj.r.squared, "\n")

# ------------------------------------------------------------------------
-
# MODELO 3 (Log-Log em área):
# ------------------------------------------------------------------------
-
modelo_v3 <- lm(log(preco_venda) ~ log(area) + quartos + vagas + suites +
                  tem_piscina + tem_churrasqueira + tem_elevador +
                  as.factor(tipologia) + as.factor(name),
                data = dados_v)

cat("\n--- Modelo Vendas 3 (Log-Log em área) ---\n")
cat("R² ajustado:", summary(modelo_v3)$adj.r.squared, "\n")

# Comparação via AIC (critério de informação de Akaike)
cat("\n--- Comparação de Modelos (AIC - menor é melhor) ---\n")
cat("Modelo 1 (Básico):", AIC(modelo_v1), "\n")
cat("Modelo 2 (Avançado):", AIC(modelo_v2), "\n")
cat("Modelo 3 (Log-Log):", AIC(modelo_v3), "\n")

# ESCOLHER O MELHOR MODELO para diagnósticos:
# (ajuste conforme o AIC - provavelmente será o modelo 3)
modelo_venda_final <- modelo_v3

# ========================================================================
# 7. DIAGNÓSTICOS ECONOMÉTRICOS - VENDAS
# ========================================================================
cat("\n=== ETAPA 7: Diagnósticos do Modelo de Vendas ===\n")

# 7a. TESTE VIF (Variance Inflation Factor) - Multicolinearidade
# Regra: VIF > 10 indica problema sério; VIF > 5 merece atenção
cat("\n--- 7a. Teste VIF (Multicolinearidade) ---\n")

#  Para calcular VIF, precisamos de um modelo SEM os fatores
(bairro/tipologia)
# porque as dummies geram VIF generalizado muito alto por construção
modelo_vif <- lm(log(preco_venda) ~ log(area) + quartos + vagas + suites +
                   tem_piscina + tem_churrasqueira + tem_elevador,
                 data = dados_v)
vif_resultado <- vif(modelo_vif)
print(vif_resultado)

# Se algum VIF > 10, remover a variável problemática
if (any(vif_resultado > 10)) {
  cat("ALERTA: Multicolinearidade detectada! Verificar variáveis com VIF >
10.\n")
```

```
} else {
  cat("OK: Nenhuma variável contínua apresenta VIF > 10.\n")
}

# VIF generalizado para fatores (GVIF)
cat("\n--- GVIF para o modelo completo ---\n")
print(vif(modelo_venda_final))

# 7b. ERROS-PADRÃO ROBUSTOS (HC1) - Correção de Heterocedasticidade

cat("\n--- 7b. Estimação com Erros-Padrão Robustos (HC1) ---\n")

# Erros robustos à la White (HC1)
coef_robustos_v <- coeftest(modelo_venda_final,
                            vcov = vcovHC(modelo_venda_final, type = "HC1"))

# Teste de Breusch-Pagan (reportar no TCC como justificativa para erros robustos)
bp_teste <- bptest(modelo_venda_final)
cat("Teste de Breusch-Pagan: estatística =", bp_teste$statistic,
    ", p-valor =", bp_teste$p.value, "\n")
if (bp_teste$p.value < 0.05) {
  cat("Heterocedasticidade confirmada (p < 0.05) — erros robustos HC1 aplicados.\n")
} else {
  cat("Homocedasticidade não rejeitada, mas erros robustos mantidos por precaução.\n")
}

# Matriz de variância-covariância robusta (guardar para usar no stargazer)
vcov_robusto_v <- vcovHC(modelo_venda_final, type = "HC1")

cat("\n--- Coeficientes com Erros Robustos (Vendas) ---\n")
print(coef_robustos_v)

# 7c. NORMALIDADE DOS RESÍDUOS
cat("\n--- 7c. Normalidade dos Resíduos ---\n")

# Shapiro-Wilk (aceita no máximo 5000 obs - usar subamostra se necessário)
set.seed(42)
amostra_residuos <- sample(residuals(modelo_venda_final),
                           min(5000, length(residuals(modelo_venda_final))))
shapiro_teste <- shapiro.test(amostra_residuos)
print(shapiro_teste)
cat("Nota: Com amostras grandes (>5000 obs), o Shapiro-Wilk tende a rejeitar\n")
cat("      normalidade mesmo com desvios pequenos. Avaliar visualmente pelo Q-Q plot.\n")

# Gráfico Q-Q dos resíduos (salvar para o TCC)
png("diagnostico_qqplot_vendas.png", width = 800, height = 600)
par(mfrow = c(1, 2))
qqnorm(residuals(modelo_venda_final), main = "Q-Q Plot - Resíduos (Vendas)")
qqline(residuals(modelo_venda_final), col = "red", lwd = 2)
hist(residuals(modelo_venda_final), breaks = 50,
     main = "Histograma dos Resíduos (Vendas)",
     xlab = "Resíduos", col = "lightblue", border = "white")
dev.off()
cat("Gráfico salvo: diagnostico_qqplot_vendas.png\n")

# 7d. REMOÇÃO DE OUTLIERS POR DISTÂNCIA DE COOK
```

```
cat("\n--- 7d. Análise de Outliers (Distância de Cook) ---\n")
cooks_v <- cooks.distance(modelo_venda_final)
limiar_cook <- 4 / nobs(modelo_venda_final)
n_outliers_cook <- sum(cooks_v > limiar_cook)
cat("Limiar de Cook (4/n):", limiar_cook, "\n")
cat("Observações influentes:", n_outliers_cook, "\n")

# Gráfico de Cook
png("diagnostico_cook_vendas.png", width = 800, height = 400)
plot(cooks_v, type = "h", main = "Distância de Cook - Vendas",
     ylab = "Distância de Cook", xlab = "Observação")
abline(h = limiar_cook, col = "red", lty = 2)
dev.off()
cat("Gráfico salvo: diagnostico_cook_vendas.png\n")

# Re-estimar SEM outliers influentes
dados_v_limpo <- dados_v[cooks_v <= limiar_cook, ]
cat("Observações após remoção por Cook:", nrow(dados_v_limpo), "\n")

modelo_venda_robusto <- lm(log(preco_venda) ~ log(area) + quartos + vagas +
suites +
                             tem_piscina + tem_churrasqueira + tem_elevador
+
                             as.factor(tipologia) + as.factor(name),
                           data = dados_v_limpo)

cat("R²               ajustado               (sem               outliers):",
summary(modelo_venda_robusto)$adj.r.squared, "\n")

# ============================================================================
# 8. ESTIMAÇÃO DO MODELO HEDÔNICO - ALUGUEL
# ============================================================================
cat("\n=== ETAPA 8: Modelo Hedônico de Aluguel ===\n")

# Modelo básico
modelo_a1 <- lm(log(preco_aluguel) ~ area + quartos + suites +
                  tem_piscina + tem_churrasqueira + tem_elevador +
                  as.factor(tipologia) + as.factor(name),
                data = dados_a)

# Modelo log-log em área
modelo_a2 <- lm(log(preco_aluguel) ~ log(area) + quartos + suites +
                  tem_piscina + tem_churrasqueira + tem_elevador +
                  as.factor(tipologia) + as.factor(name),
                data = dados_a)

cat("R² ajustado (linear):", summary(modelo_a1)$adj.r.squared, "\n")
cat("R² ajustado (log-log):", summary(modelo_a2)$adj.r.squared, "\n")
cat("AIC (linear):", AIC(modelo_a1), "\n")
cat("AIC (log-log):", AIC(modelo_a2), "\n")

modelo_aluguel_final <- modelo_a2  # ajustar conforme AIC

# ============================================================================
# 9. DIAGNÓSTICOS - ALUGUEL
# ============================================================================
cat("\n=== ETAPA 9: Diagnósticos do Modelo de Aluguel ===\n")

# VIF
modelo_vif_a <- lm(log(preco_aluguel) ~ log(area) + quartos + suites +
                     tem_piscina + tem_churrasqueira + tem_elevador,
```

```
                  data = dados_a)
cat("\n--- VIF (Aluguel) ---\n")
print(vif(modelo_vif_a))

# Erros-Padrão Robustos (HC1) - Aluguel
cat("\n--- Erros-Padrão Robustos HC1 (Aluguel) ---\n")

coef_robustos_a <- coeftest(modelo_aluguel_final,
                            vcov = vcovHC(modelo_aluguel_final, type =
"HC1"))

bp_a <- bptest(modelo_aluguel_final)
cat("Teste de Breusch-Pagan: estatística =", bp_a$statistic,
    ", p-valor =", bp_a$p.value, "\n")
if (bp_a$p.value < 0.05) {
  cat("Heterocedasticidade confirmada — erros robustos HC1 aplicados.\n")
} else {
  cat("Homocedasticidade não rejeitada, mas erros robustos mantidos por
precaução.\n")
}

vcov_robusto_a <- vcovHC(modelo_aluguel_final, type = "HC1")

cat("\n--- Coeficientes com Erros Robustos (Aluguel) ---\n")
print(coef_robustos_a)

# Q-Q plot
png("diagnostico_qqplot_aluguel.png", width = 800, height = 600)
par(mfrow = c(1, 2))
qqnorm(residuals(modelo_aluguel_final), main = "Q-Q Plot - Resíduos
(Aluguel)")
qqline(residuals(modelo_aluguel_final), col = "red", lwd = 2)
hist(residuals(modelo_aluguel_final), breaks = 50,
     main = "Histograma dos Resíduos (Aluguel)",
     xlab = "Resíduos", col = "lightblue", border = "white")
dev.off()

# Cook's distance
cat("\n--- Distância de Cook (Aluguel) ---\n")
cooks_a <- cooks.distance(modelo_aluguel_final)
limiar_cook_a <- 4 / nobs(modelo_aluguel_final)
n_outliers_cook_a <- sum(cooks_a > limiar_cook_a)
cat("Limiar de Cook (4/n):", limiar_cook_a, "\n")
cat("Observações influentes:", n_outliers_cook_a, "\n")

# Gráfico de Cook - Aluguel
png("diagnostico_cook_aluguel.png", width = 800, height = 400)
plot(cooks_a, type = "h", main = "Distância de Cook - Aluguel",
     ylab = "Distância de Cook", xlab = "Observação")
abline(h = limiar_cook_a, col = "red", lty = 2)
dev.off()
cat("Gráfico salvo: diagnostico_cook_aluguel.png\n")

dados_a_limpo <- dados_a[cooks_a <= limiar_cook_a, ]

modelo_aluguel_robusto <- lm(log(preco_aluguel) ~ log(area) + quartos +
suites +
                               tem_piscina + tem_churrasqueira +
tem_elevador +
                               as.factor(tipologia) + as.factor(name),
                             data = dados_a_limpo)
```

```
cat("R²               ajustado               (sem               outliers):",
summary(modelo_aluguel_robusto)$adj.r.squared, "\n")

# ============================================================================
# 10. EXPORTAÇÃO DE RESULTADOS PARA O ARTIGO
# ============================================================================
cat("\n=== ETAPA 10: Exportando Resultados ===\n")

# 10a. Tabela comparativa dos modelos (para o Word)
# Extrai apenas as variáveis estruturais (não as 82 dummies de bairro)
gerar_tabela <- function(modelo, nome_modelo) {
  s <- summary(modelo)
  coefs <- as.data.frame(s$coefficients)
  colnames(coefs) <- c("Coeficiente", "Erro_Padrao", "t_valor", "p_valor")

  # Significância
  coefs$Sig <- ifelse(coefs$p_valor < 0.001, "***",
                      ifelse(coefs$p_valor < 0.01, "**",
                             ifelse(coefs$p_valor < 0.05, "*",
                                    ifelse(coefs$p_valor   <   0.1,   ".",
"NS"))))

  coefs$Variavel <- rownames(coefs)
  coefs$Modelo <- nome_modelo

  # Separar variáveis estruturais e de bairro
  is_bairro <- grepl("as.factor\\(name\\)", coefs$Variavel)
  is_tipologia <- grepl("as.factor\\(tipologia\\)", coefs$Variavel)

  lista <- list(
    estruturais = coefs[!is_bairro & !is_tipologia, ],
    tipologia = coefs[is_tipologia, ],
    bairros = coefs[is_bairro, ],
    r2_ajustado = s$adj.r.squared,
    f_stat = s$fstatistic[1],
    n_obs = nobs(modelo),
    aic = AIC(modelo)
  )
  return(lista)
}

tab_venda <- gerar_tabela(modelo_venda_robusto, "Vendas")
tab_aluguel <- gerar_tabela(modelo_aluguel_robusto, "Aluguel")

# Exportar para Excel (3 abas)
write_xlsx(list(
  "Var_Estruturais_Venda" = tab_venda$estruturais,
  "Tipologia_Venda" = tab_venda$tipologia,
  "Bairros_Venda" = tab_venda$bairros,
  "Var_Estruturais_Aluguel" = tab_aluguel$estruturais,
  "Tipologia_Aluguel" = tab_aluguel$tipologia,
  "Bairros_Aluguel" = tab_aluguel$bairros
), path = "resultados_modelo_hedonico.xlsx")

cat("Tabelas salvas: resultados_modelo_hedonico.xlsx\n")

# 10b. Resumo dos modelos
cat("\n=======================================================\n")
cat("RESUMO FINAL DOS MODELOS\n")
cat("=======================================================\n")
```

```
cat("\nMODELO DE VENDAS:\n")
cat("  R² ajustado:", tab_venda$r2_ajustado, "\n")
cat("  F-statistic:", tab_venda$f_stat, "\n")
cat("  N obs:", tab_venda$n_obs, "\n")
cat("  AIC:", tab_venda$aic, "\n")

cat("\nMODELO DE ALUGUEL:\n")
cat("  R² ajustado:", tab_aluguel$r2_ajustado, "\n")
cat("  F-statistic:", tab_aluguel$f_stat, "\n")
cat("  N obs:", tab_aluguel$n_obs, "\n")
cat("  AIC:", tab_aluguel$aic, "\n")

# 10c. Tabela LaTeX/HTML via stargazer com erros robustos HC1
# Calcular erros robustos para os modelos finais (pós-Cook)
vcov_robusto_vf <- vcovHC(modelo_venda_robusto, type = "HC1")
vcov_robusto_af <- vcovHC(modelo_aluguel_robusto, type = "HC1")

# Extrair erros-padrão robustos
se_robusto_v <- sqrt(diag(vcov_robusto_vf))
se_robusto_a <- sqrt(diag(vcov_robusto_af))

stargazer(modelo_venda_robusto, modelo_aluguel_robusto,
          type = "text",
          se = list(se_robusto_v, se_robusto_a),  # erros robustos HC1
          title = "Modelos Hedônicos - Porto Alegre (2026) — Erros Robustos
HC1",
          dep.var.labels = c("log(Preço Venda)", "log(Preço Aluguel)"),
          omit = c("as.factor\\(name\\)", "as.factor\\(tipologia\\)"),
          omit.labels = c("Dummies de Bairro", "Dummies de Tipologia"),
          add.lines = list(
            c("Dummies de Bairro", "Sim", "Sim"),
            c("Dummies de Tipologia", "Sim", "Sim"),
            c("Erros-padrão", "Robustos (HC1)", "Robustos (HC1)")
          ),
          notes = "Erros-padrão robustos (HC1) entre parênteses.",
          out = "tabela_modelos_tcc.txt")

cat("\nTabela stargazer salva: tabela_modelos_tcc.txt\n")

# ============================================================================
# 11. GRÁFICOS PARA O ARTIGO
# ============================================================================

# Resíduos vs Valores Ajustados (detecta heterocedasticidade visual)
png("diagnostico_residuos_vs_ajustados.png", width = 1200, height = 500)
par(mfrow = c(1, 2))
plot(fitted(modelo_venda_robusto), residuals(modelo_venda_robusto),
     pch = 20, col = rgb(0, 0, 0, 0.1), cex = 0.5,
     main = "Resíduos vs Ajustados (Vendas)",
     xlab = "Valores Ajustados", ylab = "Resíduos")
abline(h = 0, col = "red", lty = 2)

plot(fitted(modelo_aluguel_robusto), residuals(modelo_aluguel_robusto),
     pch = 20, col = rgb(0, 0, 0, 0.1), cex = 0.5,
     main = "Resíduos vs Ajustados (Aluguel)",
     xlab = "Valores Ajustados", ylab = "Resíduos")
abline(h = 0, col = "red", lty = 2)
dev.off()

# ============================================================================
# 12. MAPAS DA AMOSTRA - PORTO ALEGRE
```

```
# ============================================================================
# Pacotes necessários (instalar apenas uma vez):
install.packages(c("sf", "geobr", "ggplot2", "viridis", "patchwork"))

cat("\n=== ETAPA 12: Gerando Mapas da Amostra ===\n")

library(sf)         # manipulação de dados espaciais
library(geobr)      # malha de municípios e bairros do Brasil (IBGE)
library(viridis)    # paleta de cores acessível
library(patchwork)  # combinar gráficos lado a lado

# 12a. Baixar geometria de Porto Alegre (bairros)
# geobr puxa direto do IBGE — precisa de internet na primeira vez
cat("Baixando malha de bairros de Porto Alegre (IBGE)...\n")

# Malha de sub-distritos / setores do município
# Opção 1: contorno do município
poa_municipio <- read_municipality(code_muni = 4314902, year = 2020)

# Opção 2: bairros (neighborhood) - disponível para capitais
poa_bairros <- tryCatch(
  read_neighborhood(year = 2010),
  error = function(e) NULL
)

# Filtrar só Porto Alegre se baixou todos os bairros
if (!is.null(poa_bairros)) {
  poa_bairros <- subset(poa_bairros, code_muni == 4314902)
  cat("Bairros carregados:", nrow(poa_bairros), "\n")
  usar_bairros <- TRUE
} else {
  cat("Bairros não disponíveis via geobr. Usando contorno do município.\n")
  usar_bairros <- FALSE
}

# 12b. Converter dados dos imóveis em objetos espaciais (sf)
pontos_venda <- st_as_sf(dados_v,
                         coords = c("longitude", "latitude"),
                         crs = 4326)  # WGS84

pontos_aluguel <- st_as_sf(dados_a,
                           coords = c("longitude", "latitude"),
                           crs = 4326)

# 12c. MAPA 1: Distribuição espacial das amostras (Venda e Aluguel)
cat("Gerando mapa de distribuição da amostra...\n")

base_map <- if (usar_bairros) {
  geom_sf(data = poa_bairros, fill = "grey95", color = "grey70", linewidth
= 0.2)
} else {
  geom_sf(data = poa_municipio, fill = "grey95", color = "grey50",
linewidth = 0.3)
}

mapa_venda <- ggplot() +
  base_map +
  geom_sf(data = pontos_venda, aes(color = log(preco_venda)),
          size = 0.4, alpha = 0.5) +
  scale_color_viridis(option = "C", name = "log(Preço)") +
  labs(title = "Imóveis à Venda",
```

```
       subtitle = paste0("n = ", nrow(dados_v), " observações")) +
  theme_minimal(base_size = 10) +
  theme(legend.position = "bottom",
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank())

mapa_aluguel <- ggplot() +
  base_map +
  geom_sf(data = pontos_aluguel, aes(color = log(preco_aluguel)),
          size = 0.4, alpha = 0.5) +
  scale_color_viridis(option = "D", name = "log(Aluguel)") +
  labs(title = "Imóveis para Aluguel",
       subtitle = paste0("n = ", nrow(dados_a), " observações")) +
  theme_minimal(base_size = 10) +
  theme(legend.position = "bottom",
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank())

# Combinar lado a lado
mapa_combinado <- mapa_venda + mapa_aluguel +
  plot_annotation(
    title = "Distribuição Espacial da Amostra — Porto Alegre (2025-2026)",
    subtitle = "Dados: ZAP Imóveis | Elaboração: Lorenzzo S. Santos",
    theme = theme(plot.title = element_text(face = "bold", size = 14))
  )

ggsave("mapa_distribuicao_amostra.png", mapa_combinado,
       width = 14, height = 8, dpi = 300)
cat("Salvo: mapa_distribuicao_amostra.png\n")

# 12d. MAPA 2: Preço médio por bairro (se houver geometria de bairros)
if (usar_bairros) {
  cat("Gerando mapa de preço médio por bairro...\n")

  # Calcular preço médio por m² por bairro
  preco_bairro_v <- aggregate(preco_m2 ~ name, data = dados_v,
                              FUN = function(x) c(media = mean(x), n =
length(x)))
  preco_bairro_v <- do.call(data.frame, preco_bairro_v)
  colnames(preco_bairro_v) <- c("name", "preco_m2_medio", "n_imoveis")

  # Juntar com geometria dos bairros (match por nome)
  # NOTA: os nomes podem não bater perfeitamente — conferir manualmente
  poa_bairros$name_lower <- tolower(poa_bairros$name_neighborhood)
  preco_bairro_v$name_lower <- tolower(preco_bairro_v$name)

  poa_com_preco <- merge(poa_bairros, preco_bairro_v,
                         by = "name_lower", all.x = TRUE)

  mapa_preco_bairro <- ggplot() +
    geom_sf(data = poa_com_preco, aes(fill = preco_m2_medio),
            color = "white", linewidth = 0.15) +
    scale_fill_viridis(option = "C", name = "R$/m²",
                       na.value = "grey90",
                       labels = scales::comma) +
    labs(title = "Preço Médio por m² de Venda por Bairro",
         subtitle = "Porto Alegre (2025-2026) | Fonte: ZAP Imóveis",
         caption = "Bairros em cinza: sem dados na amostra") +
    theme_minimal(base_size = 11) +
```

```
    theme(legend.position = "right",
          axis.text = element_blank(),
          axis.ticks = element_blank(),
          panel.grid = element_blank())

  ggsave("mapa_preco_m2_bairro.png", mapa_preco_bairro,
         width = 10, height = 10, dpi = 300)
  cat("Salvo: mapa_preco_m2_bairro.png\n")
}

# 12e. MAPA 3: Contagem de imóveis por bairro (mostra representatividade)
cat("Gerando mapa de contagem por bairro...\n")

contagem_bairro <- as.data.frame(table(dados_v$name))
colnames(contagem_bairro) <- c("name", "n_venda")

mapa_contagem <- ggplot(contagem_bairro, aes(x = reorder(name, n_venda), y
= n_venda)) +
  geom_col(fill = "steelblue", width = 0.7) +
  coord_flip() +
  labs(title = "Número de Imóveis por Bairro (Venda)",
       x = "", y = "Quantidade de imóveis na amostra") +
  theme_minimal(base_size = 8) +
  theme(axis.text.y = element_text(size = 5))

ggsave("grafico_imoveis_por_bairro.png", mapa_contagem,
       width = 8, height = 12, dpi = 300)
cat("Salvo: grafico_imoveis_por_bairro.png\n")

cat("\n=== MAPAS FINALIZADOS ===\n")

# ============================================================================

cat("\n=== SCRIPT FINALIZADO COM SUCESSO ===\n")
cat("Arquivos gerados:\n")
cat("  - resultados_modelo_hedonico.xlsx\n")
cat("  - tabela_modelos_tcc.txt\n")
cat("  - diagnostico_qqplot_vendas.png\n")
cat("  - diagnostico_qqplot_aluguel.png\n")
cat("  - diagnostico_cook_vendas.png\n")
cat("  - diagnostico_cook_aluguel.png\n")
cat("  - diagnostico_residuos_vs_ajustados.png\n")
cat("  - mapa_distribuicao_amostra.png\n")
cat("  - mapa_preco_m2_bairro.png\n")
cat("  - grafico_imoveis_por_bairro.png\n")
```