



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIX SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2025**

Modelo de regressão para preço de imóveis usados em Feira de Santana

Guilherme Ribeiro Carvalho¹; Mauricio Santana Lordelo²

1. Bolsista PROBIC/UEFS, Graduando de AGRONOMIA, e-mail: guirc2003@gmail.com

2. Orientador, DEPARTAMENTO DE CIÊNCIAS EXATAS, e-mail: mslordelo@uefs.br

PALAVRAS-CHAVE: modelagem; imóveis; regressão; *web scraping*

INTRODUÇÃO

A técnica de *web scraping* consiste em recuperar dados de páginas da *web*, seja de forma completa ou seletiva, para armazenamento em arquivos ou bancos de dados e posterior análise. Essa abordagem permite obter informações de diversas áreas do conhecimento e tem se tornado cada vez mais utilizada em pesquisas, especialmente no estudo do mercado imobiliário (SOUZA et al., 2023), que inclui diversas atividades que estão relacionadas com a construção e comercialização de imóveis, incluindo a compra de terrenos, o processo construtivo e a oferta de serviços complementares, como marketing, vendas, financiamento e manutenção após a entrega dos empreendimentos, tendo um papel importante para o desenvolvimento urbano (BOTELHO, 2005). Assim, o mercado imobiliário gera intensa atividade econômica e estimula o crescimento de serviços que vão além do próprio setor, e compreender esses fatores é fundamental para subsidiar decisões de investidores, gestores e planejadores urbanos, tornando a análise do mercado mais precisa e fundamentada. Dessa forma, o presente trabalho visa investigar os determinantes do preço de imóveis residenciais urbanos, utilizando modelos de regressão linear múltipla para avaliar a influência das variáveis observadas.

A regressão linear múltipla é uma técnica estatística amplamente utilizada para modelar a relação entre uma variável dependente contínua e duas ou mais variáveis independentes. Essa abordagem permite estimar como alterações nas variáveis explicativas impactam a variável de interesse, controlando simultaneamente os efeitos de outras variáveis (HAIR Jr. et al., 2009). É particularmente útil em contextos nos quais múltiplos fatores influenciam o fenômeno estudado, como no mercado imobiliário, onde características como área construída, localização e número de quartos podem afetar o valor de um imóvel. Além disso, a regressão linear múltipla possibilita a identificação de variáveis significativas e a quantificação de seus efeitos.

MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)

A pesquisa foi realizada com base em dados de imóveis à venda no município de Feira de Santana-BA, coletados a partir de anúncios de imóveis disponibilizados no site *Viva Real*, onde foram coletados em junho de 2025, a base de dados foi construída a partir de técnicas de *web scraping*, utilizando a biblioteca *RSelenium* no software livre

R, onde foi criado um *bot* que acessava as páginas e extraía as variáveis preço (R\$), área (m²), número de quartos, número de banheiros, número de vagas de garagem, tipo de imóvel, categorizada em três níveis: Apartamento, Casa e Casa em Condomínio e o endereço do imóvel, e os organizava em uma planilha do Excel.

Os dados passaram por etapas de limpeza e padronização, incluindo remoção de duplicatas, tratamento de valores ausentes e exclusão de *outliers* que pudessem distorcer as análises. A variável endereço foi classificada em cinco regiões administrativas do município conforme o bairro do imóvel, o preço e a área foram transformados em escalas logarítmicas. A modelagem estatística foi realizada por meio de Regressão Linear Múltipla (RLM), avaliando a significância das variáveis independentes e a qualidade do ajuste do modelo, escolha do conjunto de variáveis a serem incluídas no modelo seguiram o método de seleção de variáveis *Stepwise* com base no critério de informação de Akaike (AIC). A adequação do modelo foi avaliada por meio da análise de variância (ANOVA), testes de significância dos coeficientes, R² ajustado e diagnóstico dos resíduos.

RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)

A base de dados utilizada na pesquisa foi composta inicialmente por 2064 observações. Após a aplicação dos procedimentos de limpeza e padronização, a amostra final considerada para as análises contou com 1993 imóveis. Com base nos critérios de seleção de variáveis, o modelo que apresentou melhor ajuste foi aquele que aplicou a transformação logarítmica nas variáveis preço e área, para reduzir a assimetria da distribuição dos dados, comum em variáveis desse tipo, além de minimizar problemas de heterocedasticidade nos resíduos (NUNES et al., 2019).

O modelo final ajustado para explicar os preços dos imóveis em Feira de Santana foi especificado da seguinte forma:

$$\begin{aligned} \log(\text{preço}) = & 3,963496 + 0,563724 \log(\text{área}) + 0,053280 \cdot \text{banheiros} \\ & + 0,060948 \cdot \text{garagem} + 0,101073 \cdot \text{quartos} - 0,059696 \cdot \text{casa} \\ & + 0,055873 \cdot \text{casa em condomínio} - 0,127445 \cdot \text{região 2} \\ & - 0,105006 \cdot \text{região 3} - 0,130431 \cdot \text{região 4} - 0,026017 \cdot \text{região 5} + \varepsilon \end{aligned}$$

Esse modelo apresentou significância global ($p < 0,001$) e coeficiente de determinação ajustado (R² ajustado) de aproximadamente 0,7203, indicando que cerca de 72% da variação nos preços dos imóveis foi explicada pelas variáveis consideradas (Chein, 2019). O erro padrão residual foi de 0,2122, valor que reforça o bom ajuste do modelo, evidenciado na Tabela 1.

Tabela 1. Análise variância do modelo selecionado.

Variável	GL	SQ	QM	F	p-valor
tipo	2	43,624	21,812	484,628	< 2.2e-16
log(área)	1	151,951	151,951	3376,099	< 2.2e-16
banheiros	1	17,870	17,870	397,043	< 2.2e-16

garagem	1	6,576	6,576	146.118	< 2.2e-16
região	4	4,561	1,140	25.334	< 2.2e-16
quartos	1	6,909	6,909	153.510	< 2.2e-16
resíduos	1983	89,251	0,045		< 2.2e-16

Legenda: 1. Análise de variância (ANOVA) do modelo de regressão linear múltipla para a variável resposta $\log(\text{preço})$, apresentando os graus de liberdade (GL), soma de quadrados (SQ), quadrado médio (QM), valores de F e p-valor das variáveis independentes incluídas no modelo.

Os gráficos de diagnóstico indicam que os resíduos apresentam distribuição aleatória, variância aproximadamente constante e normalidade próxima, sem outliers ou pontos influentes relevantes, sugerindo que o modelo possui ajuste adequado e pressupostos estatísticos atendidos, podendo ser observado na Figura 1.

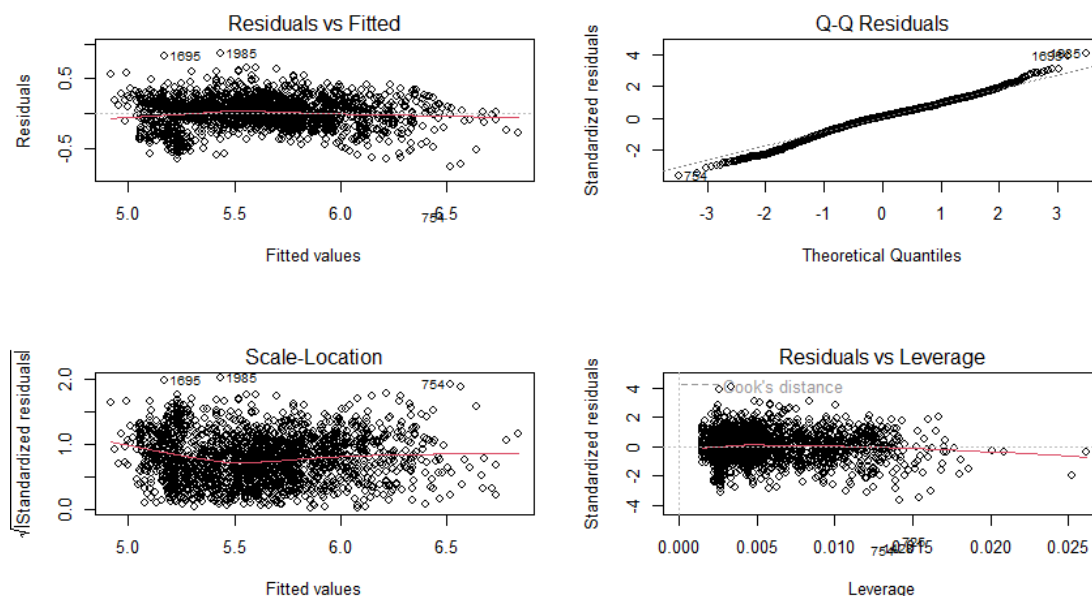


Figura 1: Gráficos de diagnóstico do modelo de regressão linear múltipla, incluindo resíduos vs valores ajustados, Q-Q plot, Scale-Location e resíduos.

CONSIDERAÇÕES FINAIS (ou Conclusão)

O estudo mostrou que características físicas, estruturais e regionais influenciam significativamente o preço de imóveis residenciais urbanos. O modelo de regressão linear múltipla apresentou boa capacidade explicativa, enquanto a coleta de dados via web scraping permitiu análises de uma quantidade grande de observações. Os resultados reforçam a importância de considerar múltiplos fatores na avaliação imobiliária.

REFERÊNCIAS

BOTELHO, A. 2005. O Financiamento e a Financeirização do Setor Imobiliário: uma análise da produção do espaço e da segregação sócio-espacial através do estudo do mercado da moradia na cidade de São Paulo. São Paulo, 2005. Tese (Doutorado) – Universidade de São Paulo, São Paulo.

SILVA, R. A.; SOUZA, J. R.; PEREIRA, M. A., 202. Potencialidades e limitações dos dados de web scraping para o mapeamento dos preços dos imóveis urbanos. *Caminhos de Geografia*, Uberlândia, v. 24, n. 2, p. 1-16.

NUNES, D. B.; PAULA BARROS NETO, J. de; FREITAS, S. M., 2019. Modelo de regressão linear múltipla para avaliação do valor de mercado de apartamentos residenciais em Fortaleza, CE. *Arquitetura & Construção*, [S.l.], v. 19, n. 3, p. 1–11. Disponível em:

<https://www.scielo.br/j/ac/a/bPkWZ5CznsJXYNHBWK78R8F/?format=pdf&lang=pt>. Acesso em: 10 set. 2025.

HAIR Jr., et al, 2009. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman. ISBN 978-85-7780-534-1.

CHEIN, F., 2019. *Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas*. Brasília: ENAP.