



LABORATÓRIO 3

LABORATORY 3

Gabriel Penha*,

Moisés Augusto†,

Carlos Siebra Boucas‡

RESUMO

Tendo em vista os gastos pessoais com moradia, especulação e outros tipos de investimentos imobiliários, é de se esperar uma crescente na literatura acerca da predição dos preços de aluguel e compra de imóveis. Dessa forma, este trabalho teve o objetivo de avaliar a relação das idades dos imóveis, de suas distâncias em jardas ao metrô mais próximo e do número de lojas de conveniência nas redondezas com o preço do imóvel por unidade de área em dólares. Os modelos ajustados não explicaram suficiente bem a variabilidade da variável resposta; mais variáveis poderiam ser consideradas, assim como um outro tipo de relação que não a linear. Dito isso, para análises mais precisas, talvez mais características sobre os imóveis sejam necessárias.

Palavras-chave: Preço de Imóveis. Regressão. Linear. Relação. Modelo.

1 INTRODUÇÃO

Tendo em vista os gastos pessoais com moradia, especulação e outros tipos de investimentos imobiliários, é de se esperar uma crescente na literatura acerca da predição dos preços de aluguel e compra de imóveis. Em seu artigo, [Morais et al. \(2015\)](#) avalia a influência da proximidade de linhas de metrô sobre o preço de imóveis no bairro de Samambaia, na cidade de Brasília. Os autores verificaram que conforme as casas ficavam mais distantes do metrô, seus valores de mercado tendiam a diminuir.

Considerando isso e a pretensão de nossa colega, coletaram-se alguns dados sobre residências nos Estados Unidos a fim de propor modelos simples, **iniciais**, para predição de preços.

Os modelos ajustados ainda não devem ser utilizados, de fato, para predizer os valores. E dois dos principais motivos são elencados a seguir:

1. Ajustaram-se modelos lineares para relações que não necessariamente eram lineares;
2. As variáveis explicativas disponíveis foram distância do metrô (como no artigo supracitado), idade da casa e número de lojas de conveniências próximas, o que deixa de fora diversos fatores importantes, por exemplo:
 - Localização do imóvel - o bairro, rua, em que ele está localizado -;
 - Tipo do imóvel - A especificação de se é apartamento ou casa faz diferença -;
 - Números de quartos e banheiros;
 - Presença ou não de mobília, entre outros.

Dito isso, as análises envolvidas devem ser utilizadas apenas para guiar os processos; recomenda-se, inclusive, uma base de dados com mais variáveis explicativas a fim de explicar a variável resposta - o preço do imóvel por unidade de área -.

Nas seguintes seções, ter-se-ão a análise exploratória dos dados, os modelos ajustados e suas interpretações (incluindo a variância dos estimadores) e algumas considerações finais.

* Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ penha.gabriel@ufba.br.

† Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ moises.augusto@ufba.br.

‡ Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ carlos.siebra@ufba.br.

2 ANÁLISE EXPLORATÓRIA

A base de dados original possuía 414 observações. Os dados já haviam sido limpos, isto é, não haviam valores faltantes, tampouco observações repetidas.

A partir daí, as estatísticas descritivas foram obtidas para as quatro variáveis: Idade do imóvel, número de lojas de conveniência a menos de uma milha do imóvel, a distância (em jardas) do imóvel ao metrô mais próximo e o preço do imóvel por unidade de área. Elas serão explicitadas abaixo, respectivamente.

A idade do imóvel possuía uma distribuição não muito simétrica, com uma leve assimetria à direita, com média de 17,7 anos e mediana de 16,1 anos. Os primeiros e terceiros quartis - respectivamente os valores que são maiores que apenas 25% e que 75% das observações - foram de 9,025 e 28,15, também respectivamente. A casa mais nova foi construída a menos de um ano, enquanto a mais velha possuía mais de 40 anos.

O número de lojas de conveniência próximas, também possuem uma distribuição levemente assimétrica, muitas casas não possuem lojas de conveniência próximas; então apesar de média e mediana serem próximas 4,094 e mediana 4, a moda é 0. A casa com mais lojas de conveniência próximas tinha 10 delas; 25% dos imóveis tinham 6 ou mais lojas de conveniência em seu entorno, e 25% tinham 1 ou menos.

Talvez uma das variáveis explicativas mais importantes dentre as analisadas, a distância do metrô para o imóvel possuía uma distribuição muito assimétrica à direita (como pode ficar evidenciado pela relação de média e mediana). 25% dos imóveis ficavam a menos de 290 jardas do metrô mais próximo; em compensação, 25% delas ficavam 1454,28, ou mais, jardas distantes. A média de distância dos metrô para com os imóveis foi de 1089,89 jardas, com mediana de 492,23.

Quanto a variável resposta, o preço médio por unidade de área foi de \$37,98, próximo dos \$38,45 da mediana. Os primeiros e terceiros quartis foram de \$27,70 e \$46,60 respectivamente. A distribuição era aproximadamente simétrica, mas contendo um valor discrepante, que foi retirado para melhorar a análise: \$117,50.

A seguir será possível visualizar algumas das relações entre as variáveis através de gráficos.

2.1 Visualização

Para um maior entendimento sobre as relações dos dados, alguns gráficos relacionando o preço dos imóveis e as variáveis explicativas foram gerados.

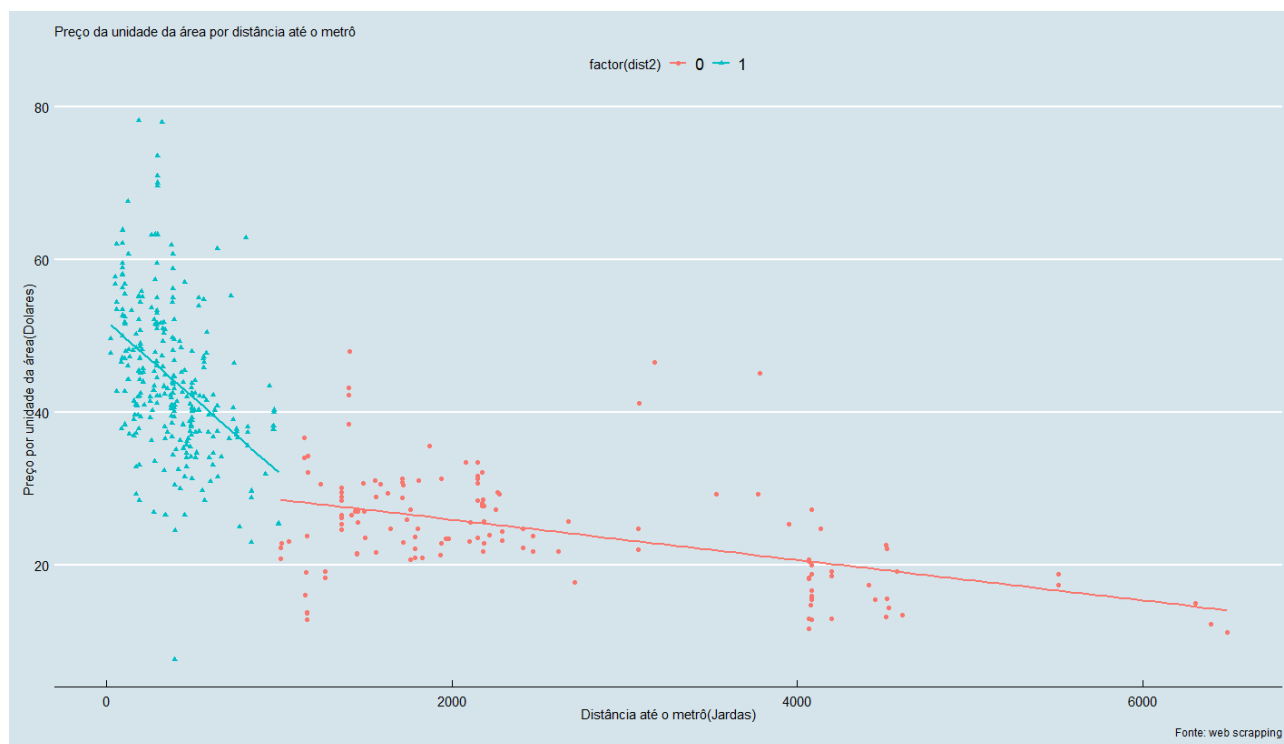


Figura 1 – Diagrama de Dispersão: Preço (\$) Vs. Distância em jardas até o metrô

Na Figura 1, separaram-se os dados em duas categorias, a primeira equivale aos imóveis que possuem uma distância

menor ou igual a 1000 jardas do metrô mais próximo. Enquanto a segunda são os imóveis com distâncias superiores a 1000 jardas. É possível visualizar uma diferença interessante nos dois grupos; desse modo, a partir de então, as análises foram feitas com essas duas distintas bases de dados.

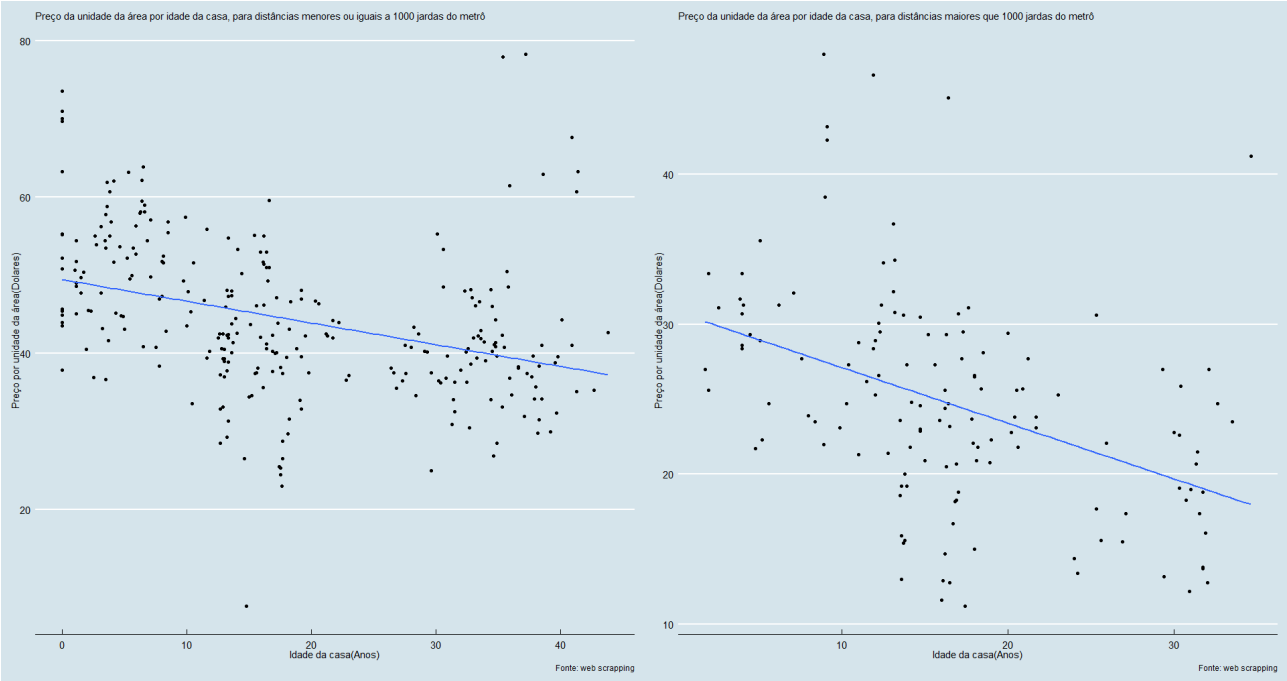


Figura 2 – Diagrama de Dispersão: Preço (\$) Vs. Idade da Casa (em anos) em ambos os grupos

Na Figura 2, percebe-se uma tendência de diminuição de preço conforme a idade da casa avança. Vale dizer, no entanto, que embora as relações pareçam lineares, elas não parecem ser muito fortes. Além disso, os dados não parecem ser homoscedásticos; isto é, a variabilidade das informações não parece ser constante.

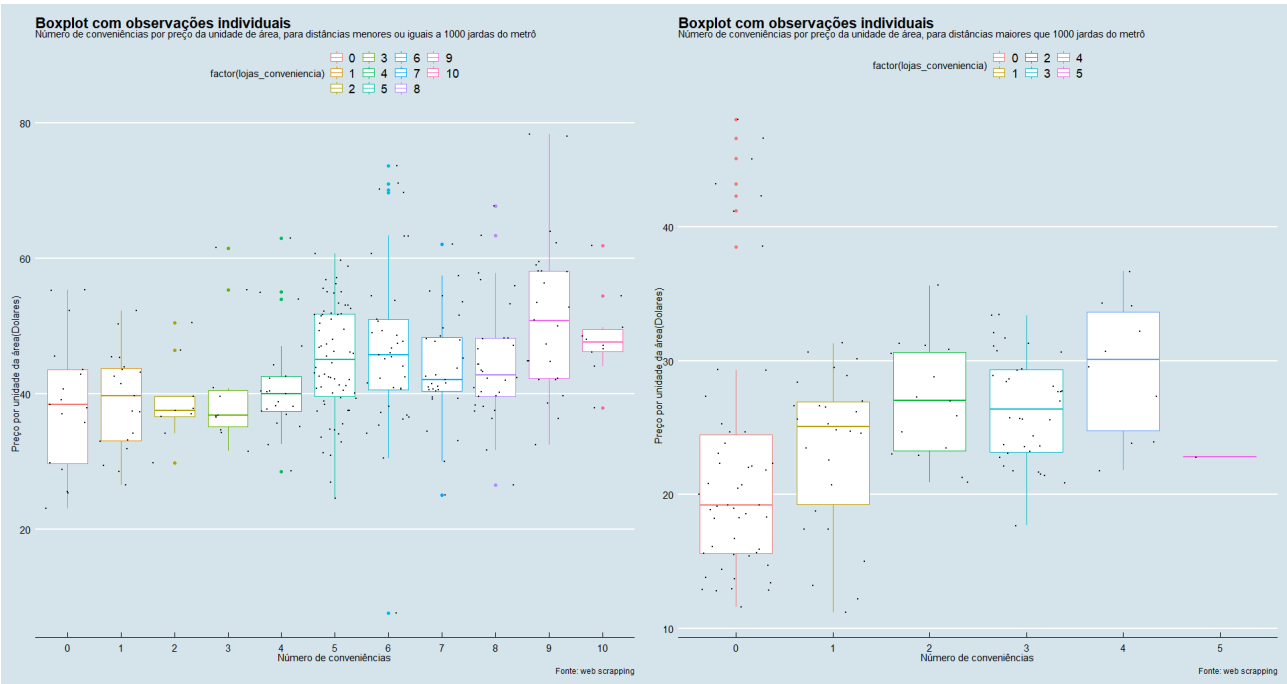


Figura 3 – Boxplots: Preço (\$) vs. Número de conveniências próximas em ambos os grupos

Na Figura 3, é possível perceber uma tendência positiva entre o preço por unidade de área e o número de conveniências próximas. A variabilidade dos preços parecia mudar a depender do número de conveniências; o que também vai contra a

suposição da homoscedasticidade.

Vale ressaltar, que para os imóveis que se situam a mais de 1000 jardas do metrô, não houveram mais que 5 lojas de conveniência. Além disso, apenas um imóvel dentre os que se situavam longe do metrô tinham 5 lojas por perto.

A partir daí, verificaram-se as correlações entre a variável resposta e as possíveis variáveis explicativas. As maiores correlações registradas foram entre o preço por unidade de área e a distância em jardas ao metrô; tanto no grupo com distância maior, quanto no com distância menor ou igual a 1000 jardas.

Assim, apesar de mais alguns modelos terem sido ajustados, apenas os que envolvem essa variável, para cada grupo, serão aqui explicitados.

3 RESULTADOS

O primeiro modelo de regressão linear simples foi ajustado no grupo com distância do imóvel ao metrô menor ou igual a 1000 jardas. A forma do modelo é visível na Equação 1.

$$Y_i = 51,8801 + -0,0198X_i + \epsilon_i \quad (1)$$

Em que Y_i , X_i e ϵ_i eram, respectivamente, o preço do imóvel por unidade de área, a distância do imóvel ao metrô mais próximo e o erro do modelo para o i -ésimo indivíduo.

A distância foi considerada uma variável significativa ao nível de 99%, ainda que explicasse menos de 20% da variabilidade de Y . A fórmula geral obtida na Equação 1 significa que cada jarda acrescida na distância do imóvel ao metrô, resultava em uma diminuição média de \$0,0198 no preço por unidade de área do imóvel. Além disso, caso o imóvel se situasse a 0 jardas do metrô, seu preço médio seria de \$51,8801. As variâncias estimadas para os parâmetros β_0 e β_1 foram respectivamente de 1,0988, 0,0025. Vale dizer que a mediana dos resíduos foi de -0,786, um valor considerado razoável.

O segundo modelo foi ajustado nas mesmas configurações do primeiro, considerando agora o grupo de casas com distâncias ao metrô maiores que 1000 jardas.

$$Y_i = 31,2363 - 0,0026X_i + \epsilon_i \quad (2)$$

Com Y_i , X_i e ϵ_i definidos como anteriormente. Este modelo indica que, para os imóveis mais distantes do metrô, cada acréscimo em jardas em tal distância reduz o preço por unidade de área do imóvel em \$-0,0026. O coeficiente de determinação ajustado foi equivalente a 0,2174. Os resíduos possuíam média próxima a 0, como esperado. As variâncias dos parâmetros β_0 e β_1 foram respectivamente 1,1907 e 0,0004.

4 CONSIDERAÇÕES FINAIS

Como mencionado anteriormente, os modelos ajustados ignoraram algumas suposições do método de regressão linear. Em um modelo linear, espera-se que os desvios sigam uma distribuição normal (simétrica) com média 0 e variância constante. No entanto, nenhum dos modelos ajustados possuía resíduos homocedásticos, o que inviabilizaria a aplicação do modelo.

Além disso, vale ressaltar que os modelos de predição explicavam muito pouco da variabilidade. Portanto, não é recomendado que sejam utilizados para predição. Implica-se em uma necessidade maior de robustez, tanto dos modelos empregados, quanto dos dados.

Vale dizer, no entanto, que assim como [Morais et al. \(2015\)](#), verificou-se que uma relação inversa entre o preço por unidade de área e a distância do imóvel ao metrô. Verificaram-se também indícios de relação entre o preço e a idade da casa (também inversa) e entre o preço e o número de lojas de conveniência próximas (direta). Contudo, não é possível afirmar que as relações são necessariamente lineares; a definição do tipo de relação que melhor descreve estes dados necessita de maiores estudos.

REFERÊNCIAS

MORAIS, A. C. de et al. Influência da proximidade à estação do metrô no valor da propriedade residencial urbana: um estudo no bairro de Samambaia em Brasília. **ANPET (esc) Aspectos Econômicos, Sociais, Políticos e Ambientais do Transporte, Ouro Preto, MG**, p. 2153–2161, 2015.

ANEXO A – FUNÇÃO - PARÂMETROS (NO R)

É possível obter o código para análises de dados no link a seguir:

<https://drive.google.com/file/d/1CUYE-p2U6r2fLLxJhjnNizqKutbQRoL/view?usp=sharing>

DECLARAÇÃO DE RESPONSABILIDADE

O(s) autor(es) é(são) o(s) único(s) responsável(eis) pelas informações contidas neste documento.