

Questão 6

Introdução

Com a intenção de determinar que variáveis influenciam o preço de venda de um imóvel uma amostra de 27 imóveis será analisada. O banco de dados em questão contém as seguintes variáveis: (i) imposto do imóvel (em 100 USD), (ii) área do terreno (em 1000 pés quadrados), (iii) área construída (em 1000 pés quadrados), (iv) idade da residência (em anos) e (v) preço de venda do imóvel (em 1000 USD). Que podem ser encontrados no site (https://www.ime.unicamp.br/~cnaber/Material_ME613_1S_2019.htm) sob o nome de “imóveis.txt”. Para tal análises será usada a metodologia dos modelos lineares homocedásticos, metodologias de verificação da qualidade do ajuste e comparação de modelos apropriado, veja Azevedo (2019). Todas as análises serão feitas com auxílio computacional do R.

Análise descritiva

Neste problema a variável respostas é o preço de venda e existe quatro variáveis explicativas que são quantitativa. Então será feito um gráfico de dispersão de cada variável explicativa pela resposta e também será analisada a correlação entre elas. Na figura 1 observa-se os quatro casos. O preço de venda parece aumentar consoante ao o aumento da área do terreno, porém não parece que uma reta modele muito bem esse aumento. Conforme a área construída cresce o preço de venda também cresce e uma reta parece modelar bem essa tendência. A relação entre o imposto e o preço aparenta ser positiva e também bem modelada por uma reta. Já a idade da residência não parece influenciar o preço de vende. Essas suposições se fortalecem na análise de correlação na figura 2, onde observa-se uma alta correlação entre idade, imposto e área construída com o preço de venda.

Análise inferencial

Considerando as características do problema, o seu objetivos e as análises descritivas, será considerado primeiramente um modelo completo sem intercepto. Já que um imóvel que não possui área, idade ou imposto em teoria não tem valor. Após o modelo completo será considerado um modelo reduzido se necessário.

Modelo completo

Seja Y_i o valor do preço de venda da i -ésima, onde $i = \{1, 2, \dots, 27\}$ tem-se o seguinte modelo:

$$Y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

onde $x_{0i}, x_{1i}, x_{2i}, x_{3i}$ são respectivamente os valores do imposto do imóvel, da área do terreno, da área construída e da idade da residência. Os parâmetros $\beta_0, \beta_1, \beta_2, \beta_3$ são respectivamente o incremento no preço de venda esperado, quando se mantém todos as outras variáveis fixas e se aumenta em uma unidade o imposto, a área do terreno, a área construída e a idade da residência. E ε_i é a componente aleatória da observação i .

Esse modelo é válido sobre as seguintes suposições: (i) $\varepsilon_i \stackrel{\text{i.i.d}}{\sim} N(0, \sigma^2)$; (ii) as observações são independentes; (iii) e a variância é constante.

Estimativa

O modelo foi ajustado segundo o método dos mínimos quadrados ordinários e a significância de seus parâmetros foi determinada a partir de um teste de hipótese utilizando as estatística t. Esse resultados se encontram na tabela 1, os parâmetros β_0 e β_2 seguem o que foi observados nas análises descritivas e foram significativos, mas os parâmetros β_1 e β_3 não foram significativos.

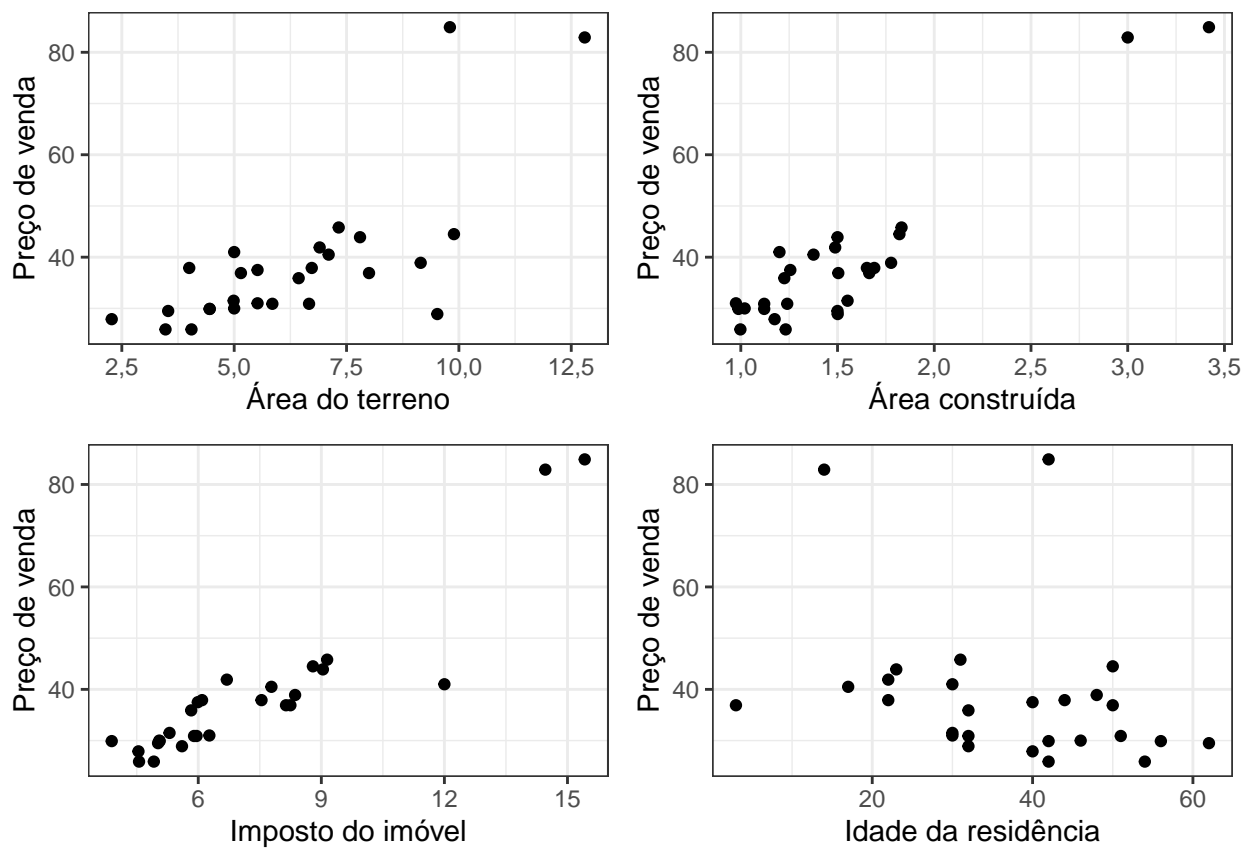


Figura 1: Comportamento do preço de venda em relação a outras variáveis.



Figura 2: Correlação entre as variáveis

Tabela 1: Estimativa para os parâmetros do modelo completo.

Termo	Estimativa	Erro Padrão	IC (95%)	Estatística t	p-valor
β_0	2,183	0,517	[1,17 ; 3,196]	4,224	< 0.001
β_1	0,357	0,455	[-0,535 ; 1,248]	0,784	0,441
β_2	13,740	2,839	[8,175 ; 19,305]	4,839	< 0.001
β_3	-0,012	0,039	[-0,089 ; 0,065]	-0,311	0,759

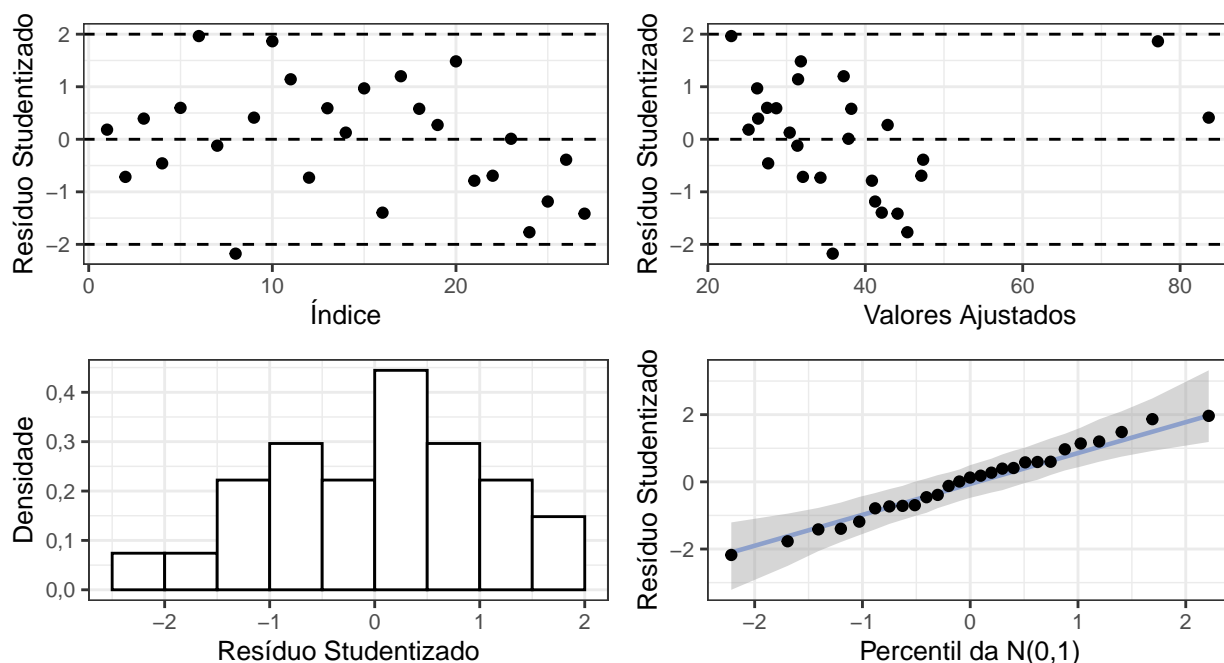


Figura 3: Diagnóstico do modelo completo.

Diagnóstico

Observando-se os gráficos de resíduos na figura 3 não há motivos para duvidar das suposições de independência dos erros e de homoscedasticidade. Já pelo histograma dos resíduos e dos gráficos quantis-quantis parece haver uma assimetria, o que é condizendo já que o preço é uma variável positiva.

Modelo reduzido

Como visto no modelo completo alguns parâmetros não foram significativos, será feito agora um modelo reduzido. Seja Y_i o valor do preço de venda da i -ésima, onde $i = \{1, 2, \dots, 27\}$ tem-se o seguinte modelo:

$$Y_i = \beta_0 x_{0i} + \beta_2 x_{2i} + \varepsilon_i$$

onde x_{0i} e x_{2i} são respectivamente os valores do imposto do imóvel, da área do terreno, da área construída e da idade da residência. Os parâmetros β_0 e β_2 são respectivamente o incremento no preço de venda esperado, quando se mantém todas as outras variáveis fixas e se aumenta em uma unidade o imposto e a área construída. E ε_i é a componente aleatória da observação i .

Tabela 2: Estimativa para os parâmetros do modelo reduzido.

Termo	Estimativa	Erro Padrão	IC (95%)	Estatística t	p-valor
β_0	2,324	0,474	[1,395 ; 3,253]	4,903	< 0.001
β_2	14,267	2,293	[9,773 ; 18,76]	6,223	< 0.001

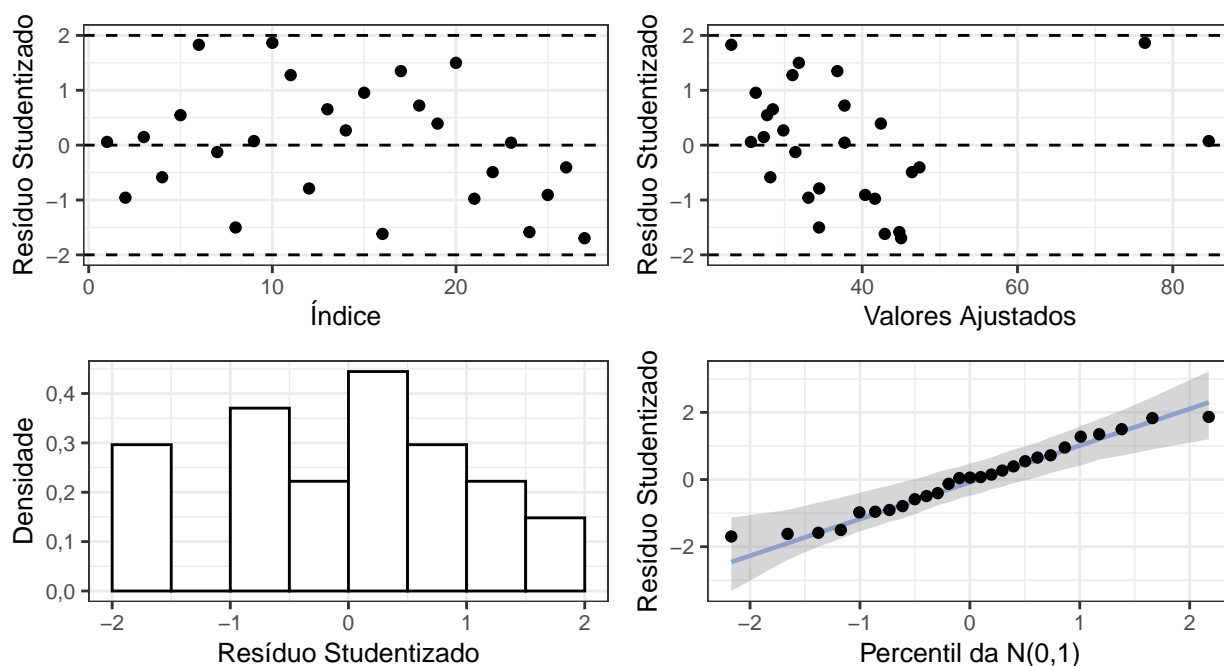


Figura 4: Diagnóstico do modelo reduzido.

Esse modelo é válido sobre as seguintes suposições: (i) $\varepsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$; (ii) as observações são independentes; (iii) e a variância é constante.

Estimativa

Os métodos de estimação são os mesmos usado no modelo completo. Observando-se a tabela 2, conclui-se que os parâmetros são significativos e fazem sentido para o modelo.

Diagnóstico

Através do gráficos de resíduos da figura 4 chegamos nas mesma conclusões que para o modelo completo, que as suposições de homocedasticidade e independência dos erros parece razoável, mas que os erros não seguem uma distribuição normal.

Comparação

Através das medidas de comparação, vistas na tabela 3 percebe-se que o modelo reduzido é melhor que o completo, já que em quase todos os casos seu valores são menores. Por isso segue-se a análise apenas com o modelo reduzido.

Tabela 3: Medidas de comparação do modelo.

	Completo	Reduzido
AIC	157,4109	154,2001
BIC	163,8900	158,0876
AICc	159,2290	154,7001
SABIC	148,1678	148,5786
HQCIC	156,9521	152,9708
-2log.lik	147,4109	148,2001

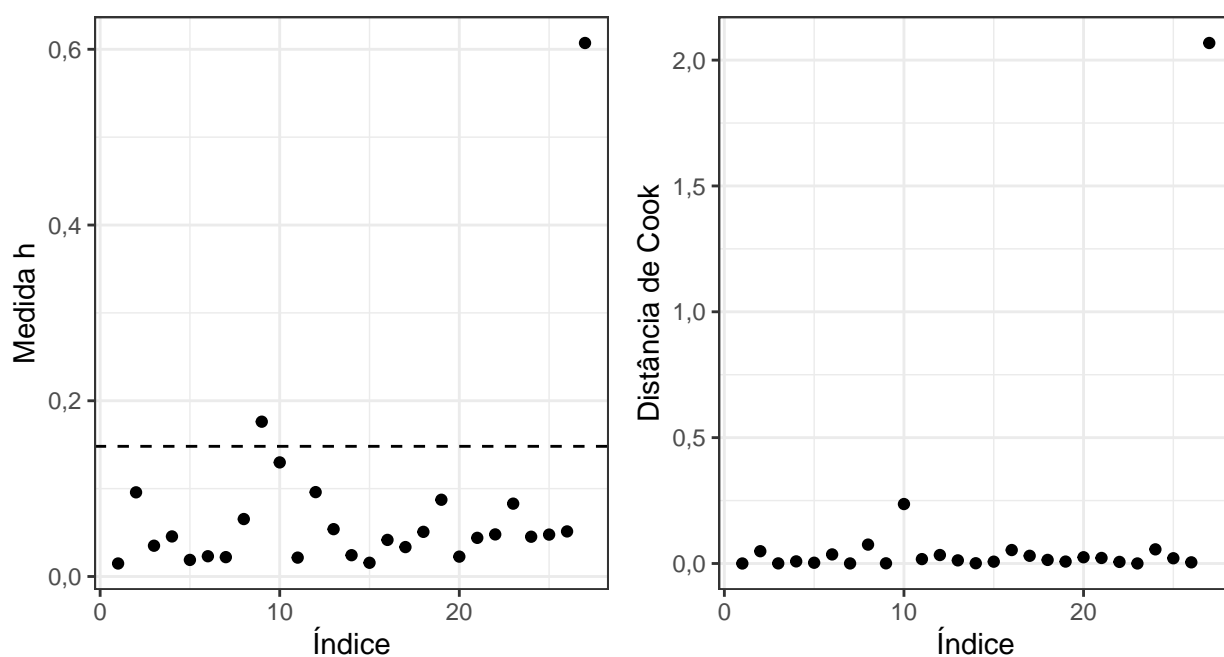


Figura 5: Análise de pontos influentes e alavancas.

Influência e Alavanca

Pelo gráficos da figura 5 percebe-se dois candidatos a pontos de influência ou alavanca, a observação 9 e a 10. O a observação 27 afeta substancialmente as estimativas, ao retirá-la do bando de dados os valores das estimativas se alteraram na ordem de 50%, mas não houve nenhuma alteração em relação as conclusões.

Conclusão

Os dois modelos propostos não se ajustaram bem aos dados, muito provavelmente por causa da não normalidade dos dados. Mas nessa classe de modelos opta-se pelo menos pior que aparenta ser o modelo reduzido. Ambos modelos sugerem que conforme o imposto pago aumenta o preço do imóvel também aumenta e que conforme a área construída aumenta o valor do imóvel também aumenta. O ponto influente provavelmente se deve ao fato dado o imposto pago o seu preço de venda é muito baixo do que deveria ser dados a reta, provavelmente algum fator não estuda influência esse caso e deve se consultar o pesquisador a respeito dele, ou ao mal ajuste do modelo.

Como o preço é uma variável positiva uma sugestão de melhorias seria usar uma distribuição positiva para modelá-lo, como Gama ou uma Normal Inversa.

Referências

Azevedo, C. L. N (2019). Notas de aula sobre Análise de regressão, http://www.ime.unicamp.br/~cnaber/Material_ME613_1S_2019.htm

Paula, G. A. (2013). Modelos de regressão com apoio computacional, versão pré-eliminar, https://www.ime.usp.br/~giapaula/texto_2013.pdf