



LABORATÓRIO 8¹ Pesquisa de Preços Imobiliários

Na discussão do laboratório 3, o diagrama de dispersão entre o preço dos imóveis e o tempo de construção do imóvel não tinha um comportamento linear. Uma possível solução para esse problema seria o uso de uma regressão polinomial. Algumas observações foram retiradas da base de dados original para a construção desse laboratório. Os dados estão disponíveis no arquivo “lab8.csv”. Considerando a relação entre a variável preço e o tempo de construção da casa, pede-se:

Objetivos

1. Uma rápida análise descritiva.
2. Um ajuste de regressão polinomial usual com grau $q = 2$ e 3 .
3. Um ajuste de regressão polinomial ortogonal com grau $q = 2$ e 3 .
4. Considere os modelos de regressão polinomial usuais com grau $q = 1, 2$ e 3 . Realizar uma comparação de modelos utilizando o princípio do resíduo condicional.
5. Considere o melhor modelo do item anterior e obtenha a estimativa do preço do imóvel para um imóvel que foi construído há 45 anos. Essa estimativa pode ser feita?
Continuando as análises com os dados sobre o preço de imóveis, pede-se
6. Ajuste um modelo de regressão múltipla com todas as covariáveis da base de dados e o polinômio escolhido no item 4. Interprete os coeficientes do modelo. Considere esse modelo para os itens a seguir.
7. Faça a seleção de modelos pelo método
 - (a) *Backward*
 - (b) *Forward*
 - (c) *Stepwise*
8. Utilize e comente os critérios de seleção de modelo
 - (a) Coeficiente de determinação
 - (b) AIC e BIC
 - (c) C_p de Mallows

Descrição das Variáveis

- *tempo*: a quanto tempo a casa foi construída, em anos;
- *metro*: distância até a estação de metrô mais próxima, em metros;
- *loja*: número de lojas de conveniência a menos de 1,6 quilômetros de distância do imóvel;
- *preco*: preço do imóvel em dólares.

¹Material iniciado pelo Discente Edre Coutinho.

Dicas:

- Regressão polinomial (cúbica): comandos
`lm(y~x+I(x^2)+I(x^3))`
`lm(y~poly(x, degree=3, raw=T))`
- Regressão polinomial (cúbica) ortogonal: comando `lm(y~poly(x, degree=3, raw=F))`
- Princípio do resíduo condicional: comando `anova(m1, m2)`, em que `m1` e `m2` são os modelos ajustados
- Seleção de modelos: comandos: `step` ou `car::subsets` ou `olsrr::ols_step_XXX`, em que `XXX` possui diversas opções, tais como: `all_possible`, `best_subset` e `forward_p`