

# Exercício 3

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:plyr':
##
##   arrange, desc, failwith, id, mutate, summarise, summarize
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

## Descrição

No arquivo **imoveis.dat** são apresentados dados relativos a uma amostra de 27 imóveis. Na ordem são apresentados os valores das seguintes variáveis:

- imposto do imóvel (em 100 USD)
- área do terreno (em 1000 pés quadrados)
- área construída (em 1000 pés quadrados)
- idade da residência (em anos)
- preço de venda do imóvel (em 1000 USD)

## Enunciado

Ajuste um modelo normal linear do preço de venda contra as demais variáveis explicativas. Use o método AIC para selecionar as variáveis explicativas. Faça uma análise de diagnóstico com o modelo selecionado. Interprete os coeficientes estimados. Seja  $y(z)$  o valor do preço de venda de um imóvel que não está na amostra com os valores das variáveis explicativas do modelo final representados por  $z$ . Como fica a estimativa intervalar de coeficiente  $(1 - \alpha)$ ,  $0 < \alpha < 1$ , para  $y(z)$ ? Alguma restrição para os valores de  $z$ ?

## Leitura dos dados

Para ler os dados no R fazemos:

```
imoveis <- data.frame(
  scan("dados/imoveis.dat", list(imposto=0, areat=0, areac=0, idade=0, preco=0)))
```

imposto	areat	areac	idade	preco
4.918	3.472	0.998	42	25.9
5.021	3.531	1.500	62	29.5

imposto	areat	areac	idade	preco
4.543	2.275	1.175	40	27.9
4.557	4.050	1.232	54	25.9
5.060	4.455	1.121	42	29.9
3.891	4.455	0.988	56	29.9

## Análise descritiva

```
df <- gather(imoveis, key = var, value = value, -preco)
ggplot(df, aes(x=value, y=preco)) + geom_point() + stat_smooth(method = "lm", se = F) +
  facet_wrap(~var, scales = "free") + xlab("Valor da variável") + ylab("Preço")
```

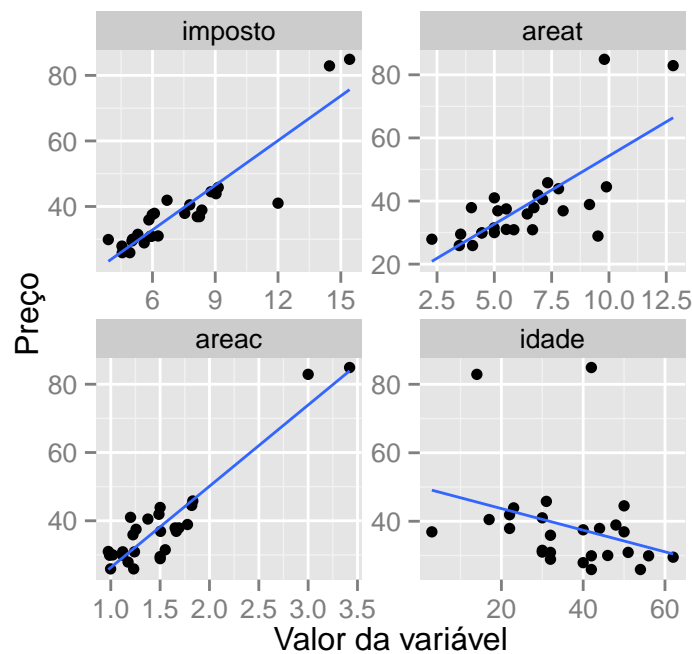


Figure 1: Gráfico de dispersão da variável resposta ‘Preço’ por todas as variáveis explicativas.

Vemos na figura 1 que todas as variáveis apresentam relação linear com a variável resposta. Quanto maior o imposto ou área de terreno ou área construída, maior a preço do imóvel. E quanto maior a idade do imóvel menor o preço do imóvel.

## Modelo

Para selecionar as variáveis explicativas vamos usar o método AIC. Neste caso como o número de variáveis explicativas é baixo, vamos ajustar todos os modelos possíveis e depois escolher aquele com o maior AIC.

Usando o código abaixo é possível ajustar todos os modelos com as variáveis do banco de dados.

```
library(meifly)
todos.modelos <- fitall(imoveis$preco, imoveis[,-5])
variaveis <- laply(todos.modelos, .fun = function(x) return(paste(names(x$coefficients)
, collapse = ", ")))
```

```
## Fitting 15 models...
```

variaveis	logL	AIC	BIC	R2	adjR2	n
(Intercept), idade	-108.28	-222.6	-226.4	0.0963	0.0601	27
(Intercept), areat, idade	-99.48	-207.0	-212.1	0.5289	0.4897	27
(Intercept), areat	-99.52	-205.0	-208.9	0.5276	0.5087	27
(Intercept), imposto, idade	-85.01	-178.0	-183.2	0.8387	0.8253	27
(Intercept), imposto, areat, idade	-83.18	-176.4	-182.8	0.8591	0.8408	27
(Intercept), imposto	-85.10	-176.2	-180.1	0.8377	0.8312	27
(Intercept), imposto, areat	-83.50	-175.0	-180.2	0.8558	0.8438	27
(Intercept), areac	-82.84	-171.7	-175.6	0.8627	0.8572	27
(Intercept), areat, areac	-81.75	-171.5	-176.7	0.8733	0.8627	27
(Intercept), areat, areac, idade	-80.19	-170.4	-176.8	0.8872	0.8725	27
(Intercept), areac, idade	-80.51	-169.0	-174.2	0.8844	0.8748	27
(Intercept), imposto, areat, areac, idade	-73.49	-159.0	-166.8	0.9313	0.9188	27
(Intercept), imposto, areat, areac	-73.75	-157.5	-164.0	0.9299	0.9208	27
(Intercept), imposto, areac, idade	-73.62	-157.2	-163.7	0.9306	0.9216	27
(Intercept), imposto, areac	-74.03	-156.1	-161.2	0.9285	0.9225	27

Vemos pela tabela acima que o modelo com maior AIC é aquele com as variáveis ‘imposto’ e ‘areac’ (área construída).