

Exercicio2.R

joaopaulodecker

2025-09-23

1 - Amostra

```
base = read.csv2("base.csv")
set.seed(07052006)
base1 = base[sample(nrow(base), 500),]
```

2 - Ajustando um modelo linear múltiplo

```
base1$Local <- as.factor(base1$Local)
modelo <- lm(Valor ~ Area + Idade + Energia + Local, data = base1)
summary(modelo)
```

##

Call:

lm(formula = Valor ~ Area + Idade + Energia + Local, data = base1)

##

Residuals:

##	Min	1Q	Median	3Q	Max
##	-32284	-6995	-173	6786	28237

##

Coefficients:

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	4098.590	3008.391	1.362	0.17370
##	Area	996.519	20.374	48.911	< 2e-16 ***
##	Idade	-504.634	44.735	-11.280	< 2e-16 ***
##	Energia	44.743	8.741	5.119	4.42e-07 ***
##	LocalB	-2478.635	1216.228	-2.038	0.04209 *
##	LocalC	-3699.276	1253.560	-2.951	0.00332 **
##	LocalD	-5833.246	1247.345	-4.677	3.77e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 9831 on 493 degrees of freedom

Multiple R-squared: 0.8427, Adjusted R-squared: 0.8408

F-statistic: 440.2 on 6 and 493 DF, p-value: < 2.2e-16

3 - Teste estatístico para verificar relação linear
`library(car)`

Loading required package: carData

`Anova(modelo)`

Anova Table (Type II tests)

##

Response: Valor

##		Sum Sq	Df	F value	Pr(>F)	
##	Area	2.3121e+11	1	2392.3288	< 2.2e-16	***
##	Idade	1.2298e+10	1	127.2484	< 2.2e-16	***
##	Energia	2.5323e+09	1	26.2014	4.417e-07	***
##	Local	2.2082e+09	3	7.6161	5.497e-05	***
##	Residuals	4.7646e+10	493			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Conclusões:

Todas as variáveis independentes são estatisticamente significativas ($p < 0.05$):

Área: $p < 2e-16$

Idade: $p < 2e-16$

Energia: $p = 4.4e-07$

Local: $p = 5.5e-05$

Logo, todas contribuem de forma significativa para explicar Valor.

4 - Interpretação dos coeficientes do modelo das variáveis que permaneceram na análise

`summary(modelo)`

##

Call:

`lm(formula = Valor ~ Area + Idade + Energia + Local, data = base1)`

##

Residuals:

##	Min	1Q	Median	3Q	Max
##	-32284	-6995	-173	6786	28237

##

Coefficients:

##		Estimate	Std. Error	t value	Pr(> t)
----	--	----------	------------	---------	----------

```
## (Intercept) 4098.590 3008.391 1.362 0.17370
## Area 996.519 20.374 48.911 < 2e-16 ***
## Idade -504.634 44.735 -11.280 < 2e-16 ***
## Energia 44.743 8.741 5.119 4.42e-07 ***
## LocalB -2478.635 1216.228 -2.038 0.04209 *
## LocalC -3699.276 1253.560 -2.951 0.00332 **
## LocalD -5833.246 1247.345 -4.677 3.77e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9831 on 493 degrees of freedom
## Multiple R-squared: 0.8427, Adjusted R-squared: 0.8408
## F-statistic: 440.2 on 6 and 493 DF, p-value: < 2.2e-16
```

Interpretações:

(Todas as variáveis possuem significância estatística)

Intercepto (4098,59): valor médio de um imóvel no bairro A, com área, idade e energia iguais a zero (sem interpretação prática relevante).

Área (996,52): cada metro quadrado adicional aumenta o preço do imóvel em aproximadamente R\$ 996,5 mil, mantendo as demais variáveis constantes.

Idade (-504,63): cada ano a mais de construção reduz o preço do imóvel em cerca de R\$ 504,6 mil, mantendo as demais variáveis constantes.

Energia (44,74): cada kWh a mais de consumo está associado a um aumento de R\$ 44,7 mil no valor do imóvel, se controladas as demais variáveis.

Imóveis no **bairro B** valem em média R\$ 2.478,6 mil a menos que os do bairro A.

Imóveis no **bairro C** valem em média R\$ 3.699,3 mil a menos que os do bairro A.

Imóveis no **bairro D** valem em média R\$ 5.833,2 mil a menos que os do bairro A.

5 - Interpretação do R^2 ajustado

```
summary(modelo)$adj.r.squared
```

```
## [1] 0.8407894
```

Isso significa que o modelo explica cerca de 84% da variabilidade no valor dos imóveis. O ajuste é considerado muito bom.

6 - Análise da importância relativa de cada variável
library(relaimpo)

Loading required package: MASS

Loading required package: boot

##

Attaching package: 'boot'

The following object is masked from 'package:car':

##

logit

Loading required package: survey

Loading required package: grid

Loading required package: Matrix

Loading required package: survival

##

Attaching package: 'survival'

The following object is masked from 'package:boot':

##

aml

##

Attaching package: 'survey'

The following object is masked from 'package:graphics':

##

dotchart

Loading required package: mitools

This is the global version of package relaimpo.

If you are a non-US user, a version with the interesting additional metric
pmvd is available

from Ulrike Groempings web site at prof.beuth-hochschule.de/groemping.

imp <- **calc.relimp**(modelo)

var.exp <- **data.frame**(**round**(**imp**\$lmg * 100, 1))

colnames(**var.exp**) <- "imp.lmg"

nome <- **rownames**(**var.exp**)

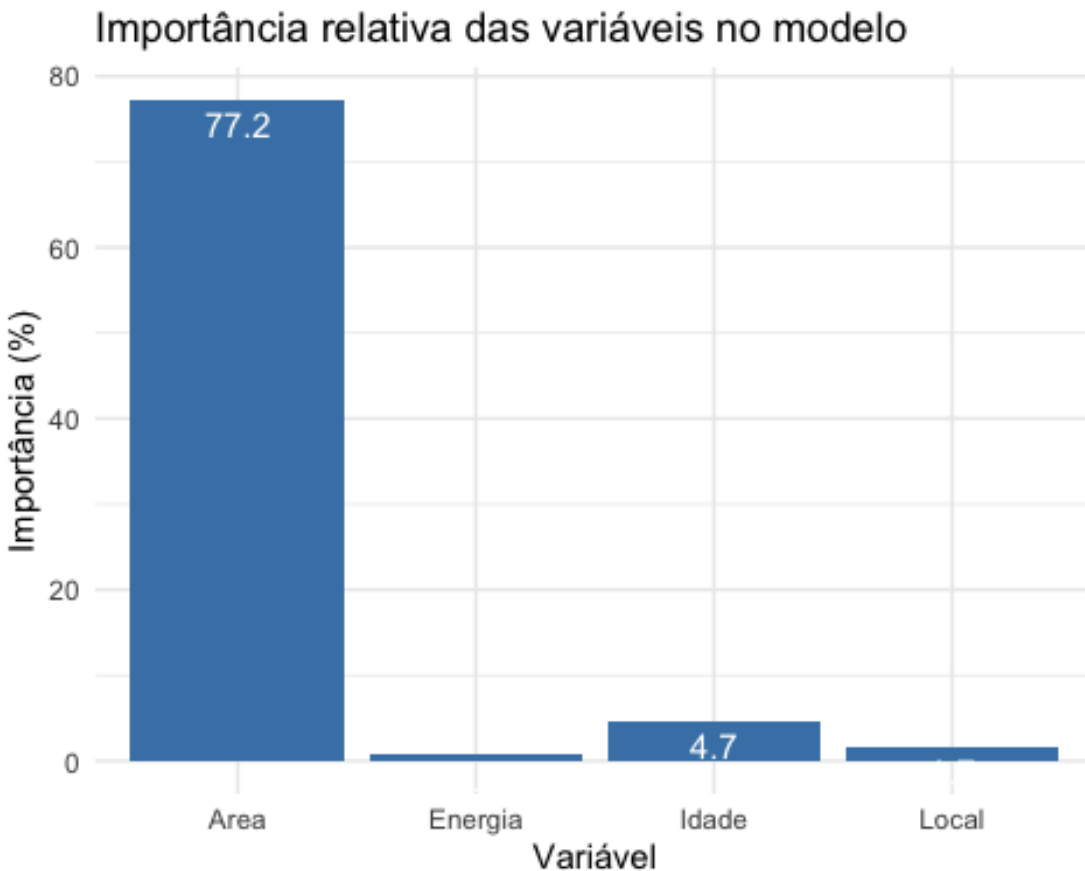
var.exp <- **data.frame**(**nome**, **var.exp**)

Gráfico

library(ggplot2)

ggplot(**var.exp**, **aes**(**x** = **nome**, **y** = **imp.lmg**)) +
 geom_bar(**stat** = "identity", **fill** = "steelblue") +

```
geom_text(aes(label = imp.lmg), vjust = 1.5, colour = "white") +
labs(x = "Variável", y = "Importância (%)",
     title = "Importância relativa das variáveis no modelo") +
theme_minimal()
```



Área foi a variável de maior importância, responsável pela maior parte da explicação da variância do preço;

Idade aparece como a segunda variável mais relevante;

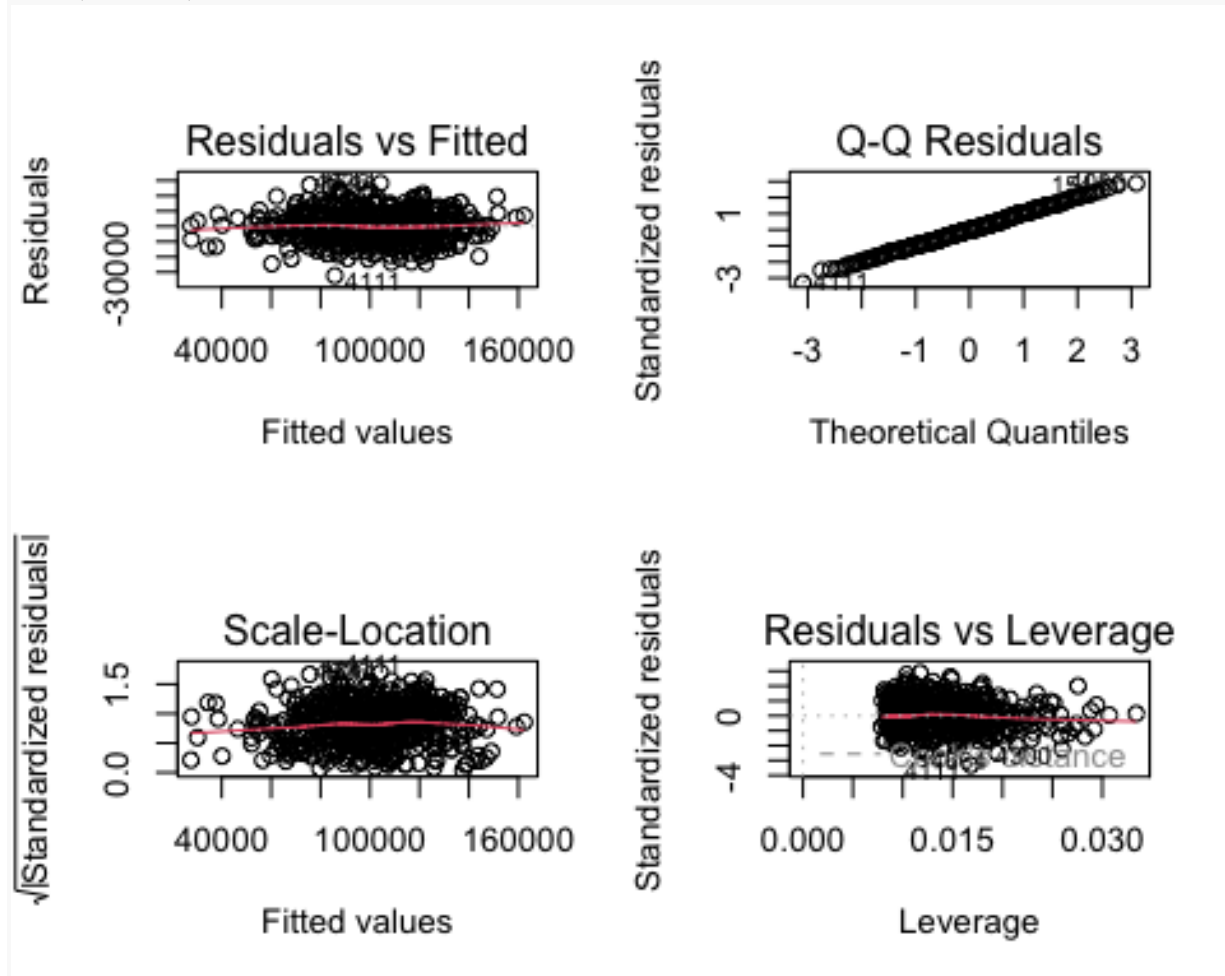
E Energia e Local também são significativas, mas com menor peso relativo.

Conclusão: a área é a principal determinante do valor dos imóveis.

7 - Análise dos resíduos do modelo (utilizando os 4 gráficos) e suas conclusões

```
par(mfrow = c(2, 2))
```

```
plot(modelo)
```



Análise dos resíduos:

Resíduos vs Valores Ajustados: os pontos estão bem espalhados, bem aleatórios, sem padrão → o modelo é linear.

QQ-Plot: os pontos seguem a linha → os resíduos têm distribuição próxima da normal.

Escala-Local: a variância dos resíduos é constante, ou seja, **não têm a mesma variância** ao longo dos valores ajustados, o que é o esperado.

Resíduos vs Leverage: não existem pontos com grande influência, ou seja, não há outliers problemáticos.

Conclusão: os resíduos não apresentam problemas e o modelo pode ser considerado **bem ajustado**