apartamento.R

Cristiane Rodrigues Maragno

2023-09-25

library(data.table)  
dados <- fread(input = paste0("apartamento.csv"), header = T, na.strings = "NA", data.table = FALSE, dec=",")  
names(dados)  
dados$Local = as.factor(dados$Local)

A)

Modelo de Regressão Múltipla

modelo <- lm(Valor ~ Area + Idade + Local, data=dados)  
summary(modelo)

##   
## Call:  
## lm(formula = Valor ~ Area + Idade + Local, data = dados)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.231 -14.403 -2.358 13.471 57.913   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.36978 11.43182 1.344 0.185845   
## Area 1.00939 0.09294 10.861 6.63e-14 \*\*\*  
## Idade -2.05789 0.51833 -3.970 0.000269 \*\*\*  
## Local2 -11.11571 7.11355 -1.563 0.125475   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22.33 on 43 degrees of freedom  
## Multiple R-squared: 0.7992, Adjusted R-squared: 0.7852   
## F-statistic: 57.06 on 3 and 43 DF, p-value: 4.873e-15

A variável Energia foi retirada do modelo por ter um p-valor maior que 0,20. Ou seja, ao nível de significancia de 20%, é muito arriscado manter essa variável

A equação desse modelo fica ao final como

Ŷ = 15,36978 + 1,00939x + -2,05789x – 11,11571x

Podemos interpretar esses resultados como

A cada m² a mais do imóvel, espera-se que aumente em R$1009,39 o valor do imóvel.

A cada ano a mais no tempo de construção do imóvel, espera-se uma diminuição de R$ 2057,89 no valor do imóvel

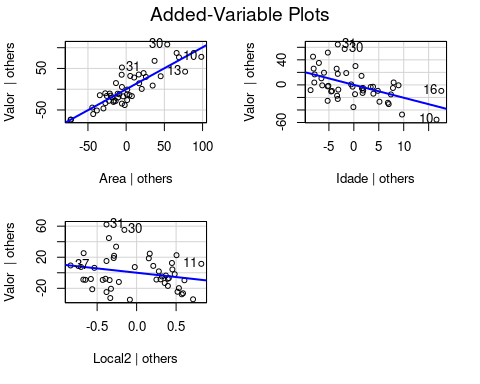
Caso o imóvel esteja na região menos valorizada, espera-se uma diminuição de R$11115,71 no valor do imóvel

B)  
library(car)

vif(modelo) # Avaliar se todas as variáveis realemente são necessárias

## Area Idade Local   
## 1.126257 1.062417 1.191844

avPlots(modelo)



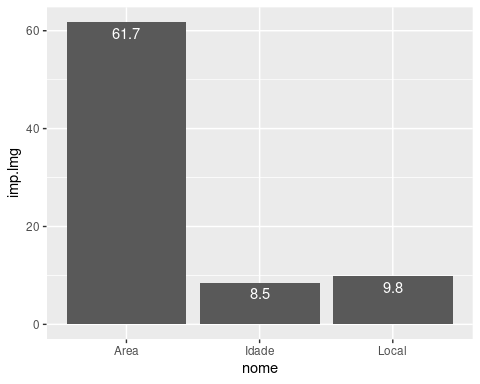
Com esse comando, podemos verificar se há uma grande relação entre as nossas variáveis, logo fazendo manter todas redundante.

Podemos concluir que todas são relevantes pois os valores são baixos, não chegando a 10.

C)  
library(relaimpo)

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)  
  
library(ggplot2)  
ggplot(var.exp,aes(nome,imp.lmg)) +  
 geom\_bar(stat = "identity")+  
 geom\_text(aes(label = imp.lmg), vjust = 1.5, lwd=6, colour = "white")

## Warning in geom\_text(aes(label = imp.lmg), vjust = 1.5, lwd = 6, colour =  
## "white"): Ignoring unknown parameters: `linewidth`



A variável da área é a mais importante porque explica 61% do valor final do imóvel, uma dirença significaticante em relação às outras.

D)

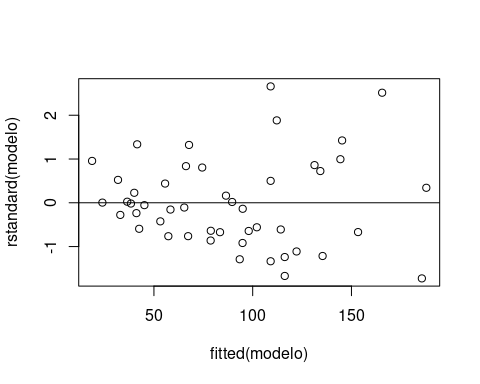
R² Ajustado  
summary(modelo)

##   
## Call:  
## lm(formula = Valor ~ Area + Idade + Local, data = dados)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.231 -14.403 -2.358 13.471 57.913   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.36978 11.43182 1.344 0.185845   
## Area 1.00939 0.09294 10.861 6.63e-14 \*\*\*  
## Idade -2.05789 0.51833 -3.970 0.000269 \*\*\*  
## Local2 -11.11571 7.11355 -1.563 0.125475   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22.33 on 43 degrees of freedom  
## Multiple R-squared: 0.7992, Adjusted R-squared: 0.7852   
## F-statistic: 57.06 on 3 and 43 DF, p-value: 4.873e-15

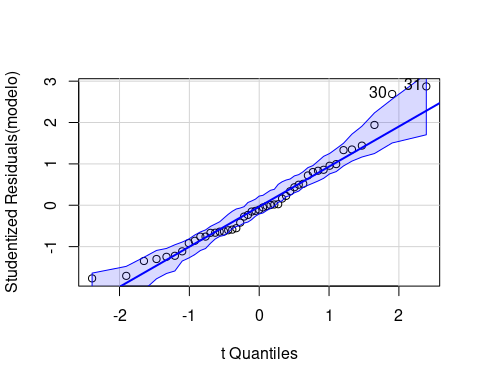
Há um R² Ajustado relevante de 0,7852, ou seja nosso modelo consegue estimar o preço do imóvel com 78% de precisão

Análise de resíduos

Os erros são aleatórios, nosso modelo está bem ajustado.  
plot(fitted(modelo), rstandard(modelo))  
abline(0,0)

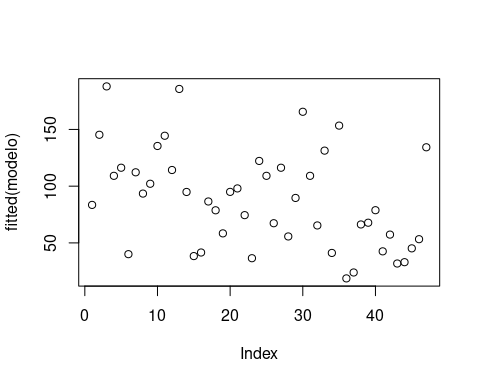


library(car)  
qqPlot(modelo)



## [1] 30 31

# Gráfico de quantil  
plot(fitted(modelo))  
abline(0,0)



library(car)  
qqPlot(modelo)

