

Exercicio3.R

joaopaulodecker

2025-10-06

```
# a- Amostra
base = read.csv2("carros.csv")
set.seed(07052006)
base1 = base[sample(nrow(base), 800),]

# b- Criação da variável idade
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

base1 <- base1 |>
  mutate(idade = 2025 - ano)

# c- Criação do modelo Linear
base1 <- base1 |>
  mutate(estado_conservacao = factor(estado_conservacao))

modelo_interacao <- lm(preco ~ idade * estado_conservacao, data = base1)

summary(modelo_interacao)

##
## Call:
## lm(formula = preco ~ idade * estado_conservacao, data = base1)
##
## Residuals:
##   Min       1Q   Median       3Q      Max
```

```
## -25436 -8344 -1231 7171 36431
##
## Coefficients:
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34495.57 1328.68 25.962 < 2e-16
***
## idade -485.83 110.20 -4.409 1.18e-05
***
## estado_conservacaoExcelente 2901.67 2073.58 1.399 0.16210
## estado_conservacaoRegular -6205.37 2382.54 -2.605 0.00937 **
## estado_conservacaoRuim -12829.12 3082.65 -4.162 3.50e-05
***
## idade:estado_conservacaoExcelente 93.66 174.04 0.538 0.59060
## idade:estado_conservacaoRegular 129.68 188.75 0.687 0.49224
## idade:estado_conservacaoRuim 300.10 244.19 1.229 0.21946
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11420 on 792 degrees of freedom
## Multiple R-squared: 0.1581, Adjusted R-squared: 0.1506
## F-statistic: 21.25 on 7 and 792 DF, p-value: < 2.2e-16

# d- Teste de interação
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
## recode

Anova(modelo_interacao)

## Anova Table (Type II tests)
##
## Response: preco
## Sum Sq Df F value Pr(>F)
## idade 4.2265e+09 1 32.3819 1.783e-08 ***
## estado_conservacao 1.3808e+10 3 35.2632 < 2.2e-16 ***
## idade:estado_conservacao 2.1646e+08 3 0.5528 0.6464
## Residuals 1.0337e+11 792
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

modelo_sem_interacao <- lm(preco ~ idade + estado_conservacao, data = base1)
summary(modelo_sem_interacao)

##
## Call:
```

```
## lm(formula = preco ~ idade + estado_conservacao, data = base1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25485  -8430  -1326   7149  36409
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    33597.12     976.79   34.396 < 2e-16 ***
## idade          -401.28      70.46   -5.695 1.73e-08 ***
## estado_conservacaoExcelente  3894.69     985.14    3.953 8.39e-05 ***
## estado_conservacaoRegular -4787.97    1096.26   -4.368 1.42e-05 ***
## estado_conservacaoRuim    -9482.53    1415.97   -6.697 4.03e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11410 on 795 degrees of freedom
## Multiple R-squared:  0.1563, Adjusted R-squared:  0.1521
## F-statistic: 36.83 on 4 and 795 DF,  p-value: < 2.2e-16
```

Explicação:

O termo de interação idade:estado_conservacao não é estatisticamente significativo (pvalor = 0.6464).

Assim, foi ajustado um novo modelo sem o termo de interação.

e- Análise de desvalorização

Observa-se que:

Carros em excelente estado possuem, em média, R\$ 3.895 a mais no preço de venda do que veículos em bom estado (categoria de referência);

Veículos em estado regular apresentam uma redução média de R\$ 4.788;

Carros em estado ruim chegam a valer cerca de R\$ 9.483 a menos.

Portanto, ao considerar a interação significativa (mesmo que não seja o caso), carros em estado ruim tendem a desvalorizar menos, uma vez que o seu coeficiente angular é menor. Já carros em excelente estado desvalorizam mais ao longo do tempo, visto que seu coeficiente é maior. (inclinação da reta).

```

# f- Gráfico para representar interação do modelo
library(ggplot2)
library(dplyr)

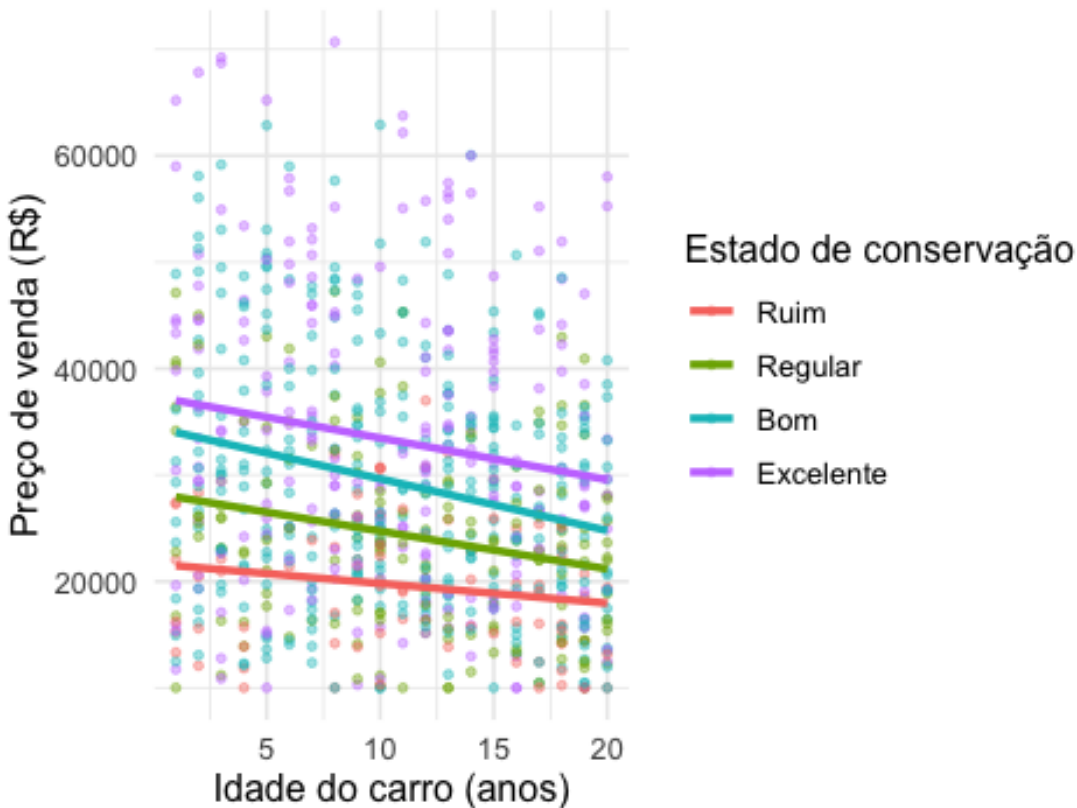
# Garantir ordem
base1 <- base1 |>
  mutate(
    estado_conservacao = factor(
      estado_conservacao,
      levels = c("Ruim", "Regular", "Bom", "Excelente")
    )
  )

ggplot(base1, aes(x = idade, y = preco, color = estado_conservacao)) +
  geom_point(alpha = 0.4, size = 1) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 1.2) +
  labs(
    title = "Relação entre Idade e Preço dos Carros por Estado de
Conservação",
    x = "Idade do carro (anos)",
    y = "Preço de venda (R$)",
    color = "Estado de conservação"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    legend.position = "right"
  )

## `geom_smooth()` using formula = 'y ~ x'

```

Idade e Preço dos Carros por Estado de Conservação



g- Intervalo de confiança e de predição

```
library(ggplot2)
```

```
library(dplyr)
```

Intervalos de confiança e predição para os dados da base

```
IC1 <- predict(modelo_sem_interacao, interval = "confidence", level = 0.95)
```

```
IC2 <- predict(modelo_sem_interacao, interval = "predict", level = 0.95)
```

```
## Warning in predict.lm(modelo_sem_interacao, interval = "predict", level = 0.95): predictions on current data refer to _future_ responses
```

Juntar resultados na mesma base

```
new <- data.frame(base1, IC1, IC2)
```

Filtrar apenas carros em excelente estado (para o gráfico ilustrativo)

```
new_excelente <- new |>
```

```
  filter(estado_conservacao == "Excelente")
```

Gráfico dos intervalos

```
ggplot(new_excelente, aes(x = idade, y = preco)) +
```

```
  geom_point(alpha = 0.4) +
```

```
  geom_line(aes(y = fit), color = "blue", linewidth = 1) +
```

Linha

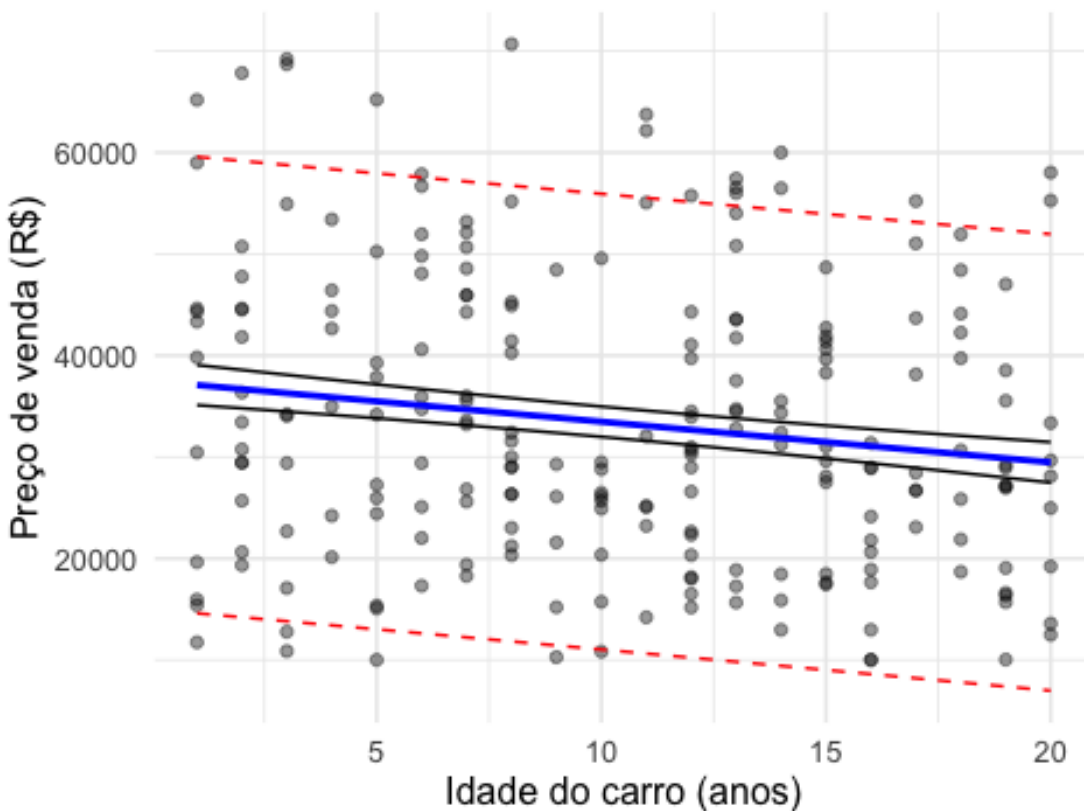
ajustada

```

    geom_line(aes(y = lwr), color = "black") + # IC
inferior
    geom_line(aes(y = upr), color = "black") + # IC
superior
    geom_line(aes(y = lwr.1), color = "red", linetype = "dashed") + # IP
inferior
    geom_line(aes(y = upr.1), color = "red", linetype = "dashed") + # IP
superior
    labs(
      title = "Intervalos de Confiança (preto) e Predição (vermelho) – Estado
Excelente",
      x = "Idade do carro (anos)",
      y = "Preço de venda (R$)"
    ) +
    theme_minimal(base_size = 12) +
    theme(plot.title = element_text(hjust = 0.5, face = "bold"))

```

os de Confiança (preto) e Predição (vermelho) — Est



```

# Previsão para um carro de 5 anos e estado 'Excelente'
novo_carro <- data.frame(idade = 5, estado_conservacao = "Excelente")

# Intervalo de confiança (média esperada)
predict(modelo_sem_interacao, novo_carro, interval = "confidence", level =
0.95)

```

```
##          fit          lwr          upr
## 1 35485.43 33819.22 37151.63

# Intervalo de predição (carro individual)
predict(modelo_sem_interacao, novo_carro, interval = "predict", level = 0.95)

##          fit          lwr          upr
## 1 35485.43 13016.54 57954.32
```

Interpretação:

O preço médio esperado de um carro com 5 anos e estado excelente é de R\$ 35.485.

O intervalo de confiança (95%) indica que a média populacional dos preços para carros com essas características está entre R\$ 33.819 e R\$ 37.152.

O intervalo de predição (95%) é mais amplo (R\$ 13.017 a R\$ 57.954) porque reflete a variabilidade natural dos preços de carros individuais.

Conclusão:

O intervalo de confiança refere-se à estimativa da média esperada do preço para carros com essas características. Já o intervalo de predição reflete a faixa de valores para o preço de um carro específico com essas características, incluindo a variação de erros individuais.