

# Exercicio1.R

joaopaulodecker

2025-09-01

*# 1 - Amostra*

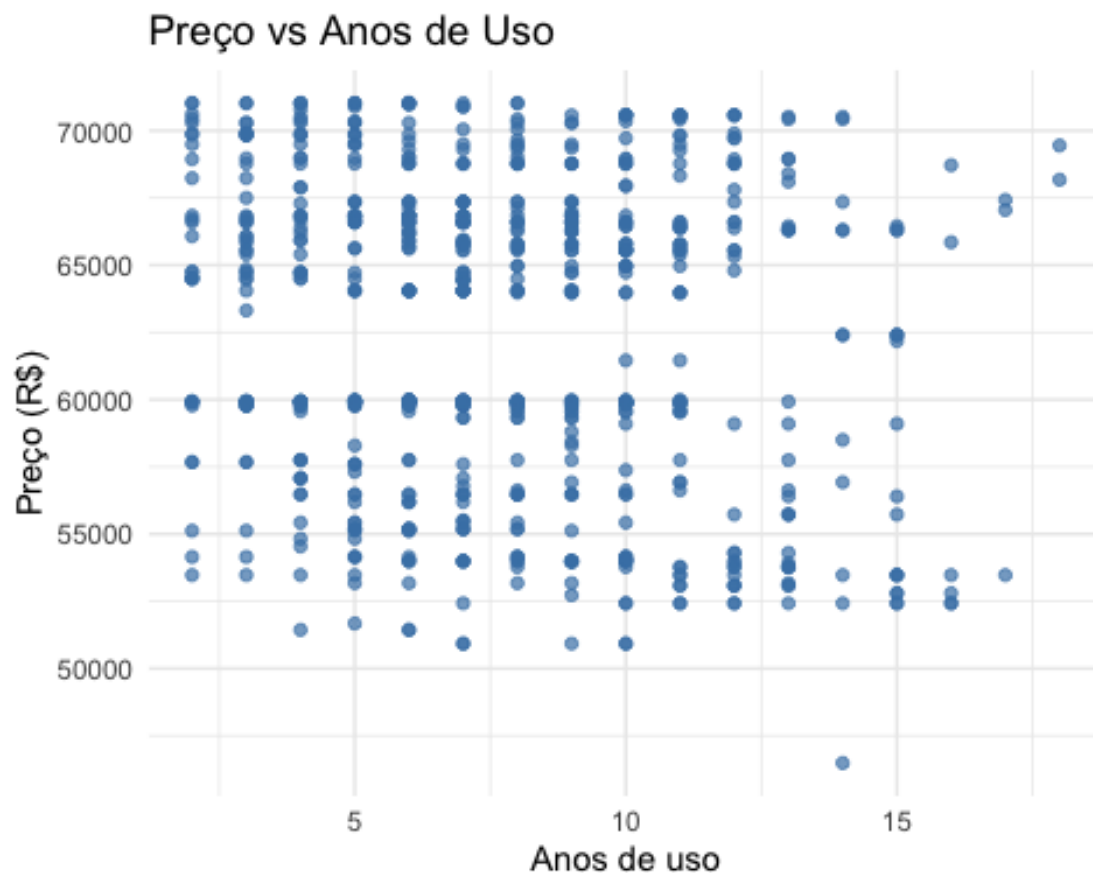
```
base = read.csv2("exercicio.csv")
set.seed(07052006)
base1 = base[sample(nrow(base), 800),]
```

*# 2 - Gráfico de dispersão de cada variável relacionada com o preço*

```
library(ggplot2)
```

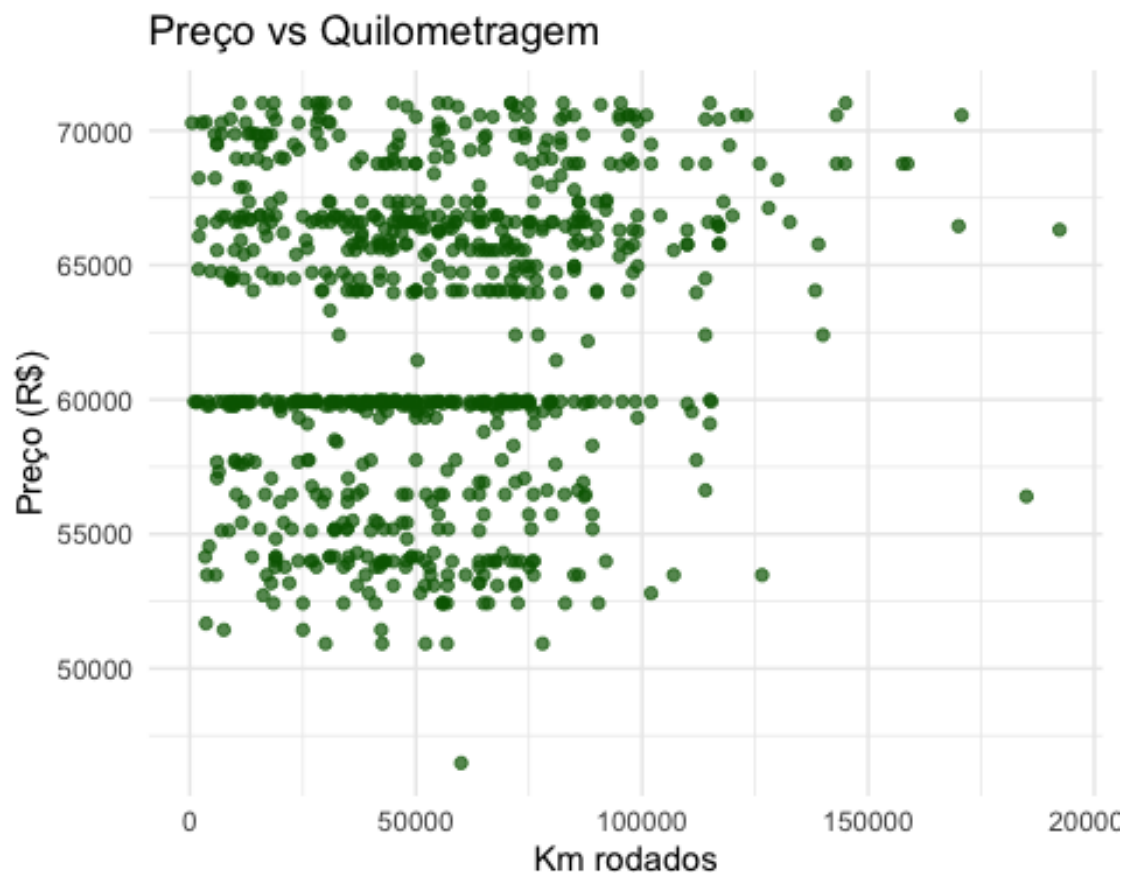
*# Gráfico 1: preço vs anos*

```
ggplot(base1, aes(x = anos, y = preco)) +
  geom_point(color = "steelblue", alpha = 0.7) +
  labs(
    title = "Preço vs Anos de Uso",
    x = "Anos de uso",
    y = "Preço (R$)"
  ) +
  theme_minimal()
```

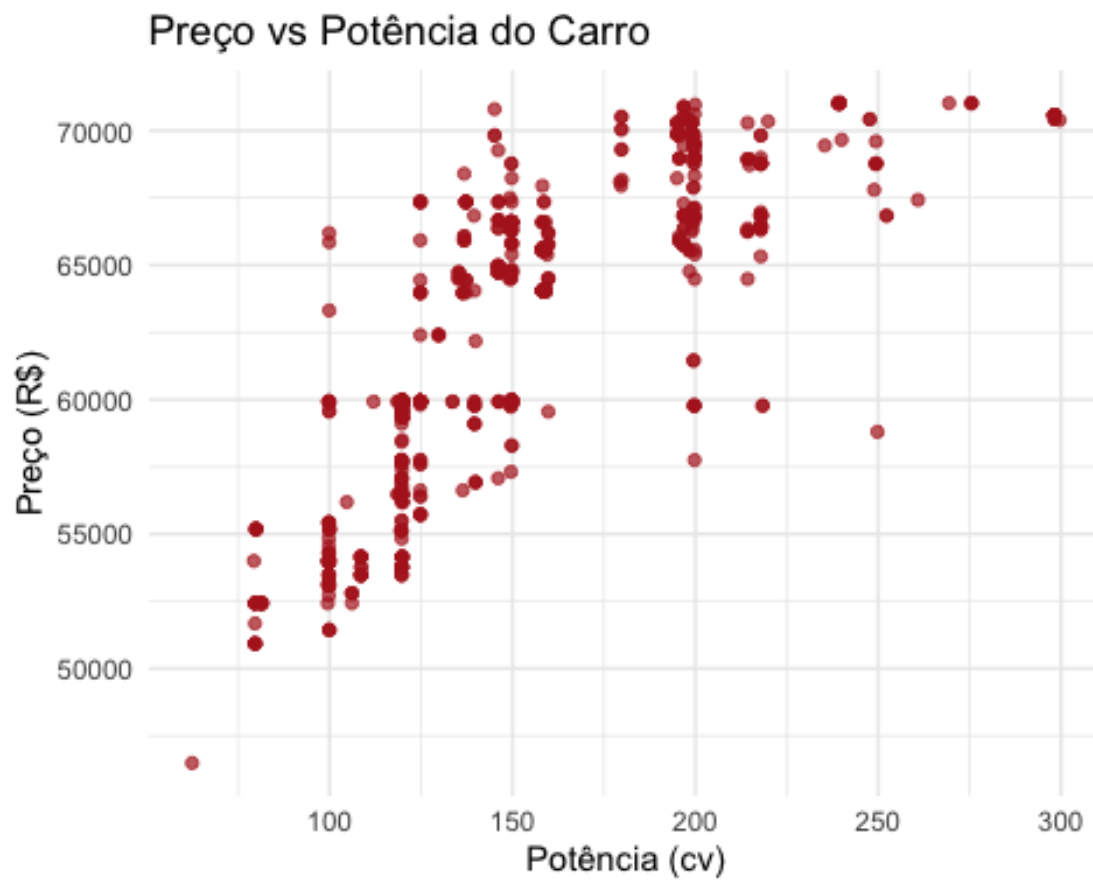


# Gráfico 2: preço vs km

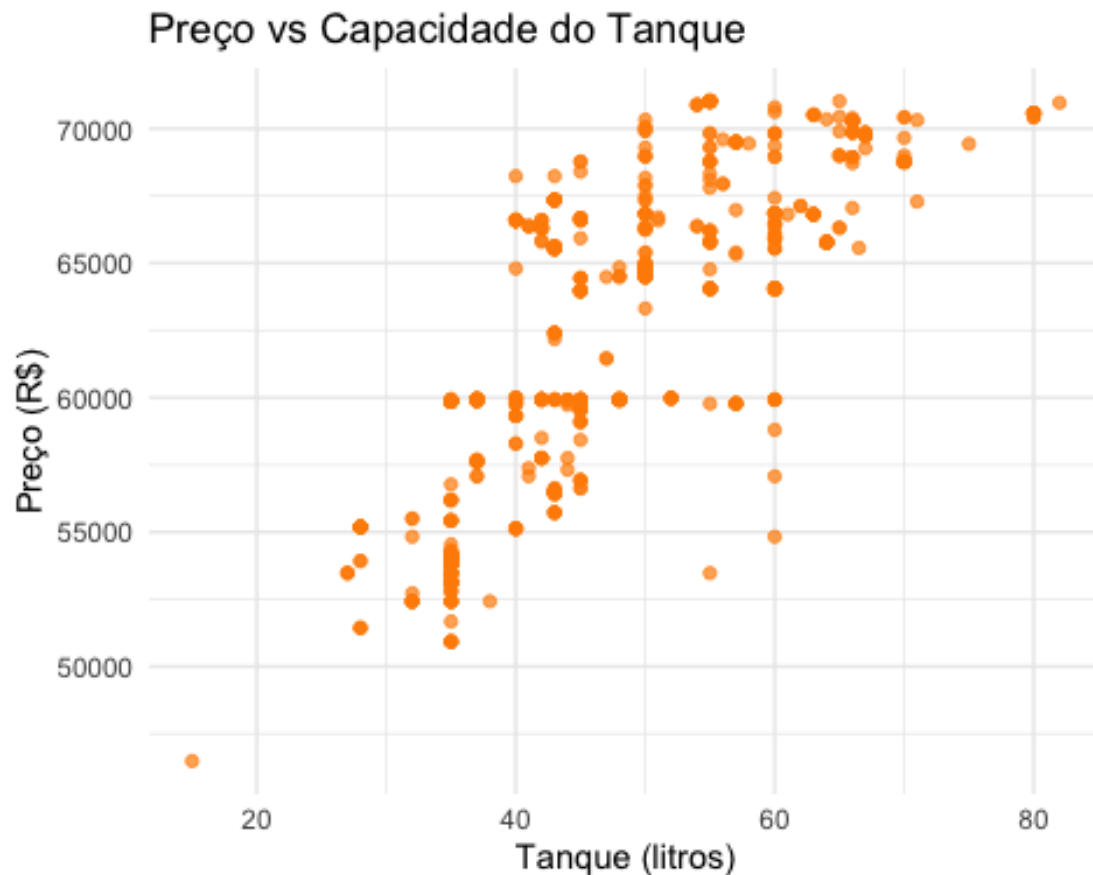
```
ggplot(base1, aes(x = km, y = preco)) +
  geom_point(color = "darkgreen", alpha = 0.7) +
  labs(
    title = "Preço vs Quilometragem",
    x = "Km rodados",
    y = "Preço (R$)"
  ) +
  theme_minimal()
```



```
# Gráfico 3: preço vs potencia
ggplot(base1, aes(x = potencia, y = preco)) +
  geom_point(color = "firebrick", alpha = 0.7) +
  labs(
    title = "Preço vs Potência do Carro",
    x = "Potência (cv)",
    y = "Preço (R$)"
  ) +
  theme_minimal()
```



```
# Gráfico 4: preço vs tanque
ggplot(base1, aes(x = tanque, y = preco)) +
  geom_point(color = "darkorange", alpha = 0.7) +
  labs(
    title = "Preço vs Capacidade do Tanque",
    x = "Tanque (litros)",
    y = "Preço (R$)"
  ) +
  theme_minimal()
```



### # 3 - Cálculo do coeficiente de correlação de cada variável quantitativa

```
vars_quant <- base1[, c("preco", "anos", "km", "potencia", "tanque")]
correlacoes <- cor(vars_quant, use = "complete.obs")
correlacoes["preco", -1]
```

```
##      anos      km  potencia  tanque
## -0.1427165  0.1460225  0.8076732  0.7540899
```

Interpretação:

“anos” -> correlação de -0.143, ou seja, fraca e negativa: mesmo que nao  
significante, quanto mais anos de uso, menor o preço, como esperado.

“km” -> correlação de 0.146, ou seja, fraca e positiva: o coeficiente é  
insignificante e diz que quanto mais km rodados, ligeiramente maior o preço,  
o que é contraintuitivo.

“potencia” -> correlação de 0.808, ou seja, forte e positiva: carros mais  
potentes tendem a ter preços mais altos, o que é coerente e indica que a  
variável tem correlação linear.

“tanque” -> correlação de 0.754, ou seja, forte e positiva: carros com maior  
capacidade de tanque também tendem a ser mais caros. Possível correlação.

#### # 4 - Intervalo de confiança da correlação entre potência e preço

```
r <- cor(base1$preco, base1$potencia, use = "complete.obs")
n <- sum(complete.cases(base1$preco, base1$potencia))
```

##### # Transformação de Fisher

```
z <- 0.5 * log((1 + r) / (1 - r))
se <- 1 / sqrt(n - 3)
```

##### # Intervalo de confiança em z

```
z_low <- z - 1.96 * se
z_high <- z + 1.96 * se
```

##### # Voltar para escala de correlação

```
r_low <- (exp(2 * z_low) - 1) / (exp(2 * z_low) + 1)
r_high <- (exp(2 * z_high) - 1) / (exp(2 * z_high) + 1)
```

##### # Resultado

```
c("inferior" = r_low, "superior" = r_high)
```

```
## inferior superior
## 0.7821456 0.8304941
```

##### Interpretação:

Utilizando a transformação de Fisher, foi calculado um intervalo de confiança de 95% para essa correlação populacional. O resultado obtido foi:

IC95% = [0,7821 ; 0,8305]

Esse intervalo sugere, com 95% de confiança, que a verdadeira correlação entre potência e preço na população se encontra nesse intervalo. Portanto, conclui-se que há uma forte associação positiva e estatisticamente significativa entre a potência do veículo e seu preço, o que reforça a relevância da variável potência como preditora no contexto analisado.

### # 5 - Modelo de regressão da variável preço x potência

```
modelo <- lm(preco ~ potencia, data = base1)
```

```
summary(modelo)
```

```
##
```

```
## Call:
```

```
## lm(formula = preco ~ potencia, data = base1)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -13017.3  -2318.8    171.6   2573.1   9087.4
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 47708.772    398.069   119.9  <2e-16 ***
```

```
## potencia      96.511       2.494    38.7  <2e-16 ***
```

```
## ---
```

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

```
##
```

```
## Residual standard error: 3287 on 798 degrees of freedom
```

```
## Multiple R-squared:  0.6523, Adjusted R-squared:  0.6519
```

```
## F-statistic: 1497 on 1 and 798 DF, p-value: < 2.2e-16
```

### # a - o coeficiente de determinação e interprete o valor.

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

```
## [1] 0.6523361
```

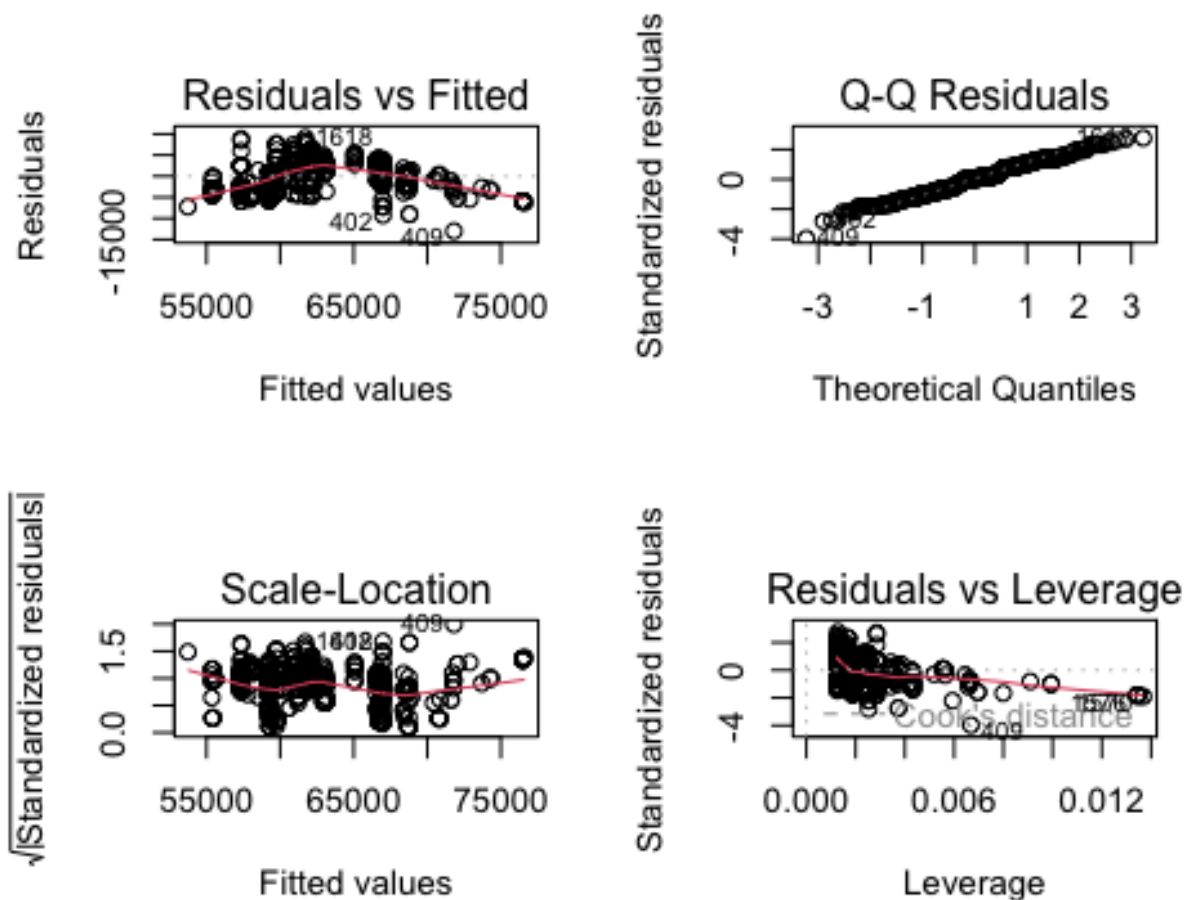
### Interpretação:

Esse valor indica que aproximadamente 65,23% da variação nos preços dos veículos pode ser explicada pela potência do motor. Isso significa que a variável potência possui alto poder explicativo sobre o comportamento da variável preço, sendo uma forte preditora no contexto analisado. Os 34,77% restantes da variabilidade nos preços são atribuídos a outros fatores não incluídos no modelo e a variação aleatória, etc.

### # b - a análise de resíduos

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

```
plot(modelo)
```



#### Interpretação:

Os pontos estão distribuídos de forma aproximadamente aleatória ao redor do zero, sem apresentar um padrão claro. Isso sugere que a suposição de linearidade e homocedasticidade (variância constante dos erros) é, em geral, atendida. E também os resíduos seguem aproximadamente a linha de referência, indicando que a suposição de normalidade dos erros é razoável.

Contudo, apesar do bom desempenho geral, os resíduos ainda apresentam uma dispersão relativamente alta. Isso indica que o modelo, embora estatisticamente adequado, pode não capturar toda a complexidade da relação entre as variáveis.



*# c - teste de significancia do modelo*

```
summary(modelo)$coefficients
```

```
##              Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 47708.7720 398.068563 119.85064 0.000000e+00
## potencia    96.5114    2.494143  38.69522 2.923956e-185
```

*Interpretação:*

O teste de significância global do modelo indicou um resultado altamente significativo. Isso demonstra que o modelo como um todo é estatisticamente significativo, ou seja, a variável potência contribui significativamente para explicar a variação no preço dos veículos.

No entanto, é importante observar que o modelo inclui apenas uma variável explicativa. A análise indica que existem outros fatores relevantes que não foram considerados e que podem melhorar a explicação da variabilidade do preço. Portanto, outros modelos poderiam proporcionar um ajuste mais robusto e explicações mais completas dos dados.

*# d - interpretação dos coeficientes do modelo*

```
coef(modelo)
```

```
## (Intercept)      potencia
## 47708.7720      96.5114
```

*Interpretação:*

O modelo ajustado foi:

$\text{preço} = 47708,77 + 96,51 \times \text{potencia}$

O intercepto (47708,77) representa o preço estimado de um carro com zero de potência, o que não tem interpretação prática, mas é necessário no contexto do modelo. O coeficiente da variável potência (96,51) indica que, para cada aumento de 1 cv de potência, o preço do veículo aumenta, em média, R\$ 96,51, mantendo os demais fatores constantes.

*# 6 - Gráfico de dispersão preço x potencia diferenciando transmissão automática e manual*

```
library(ggplot2)
```

```
ggplot(base1, aes(x = potencia, y = preco, color = transm)) +
  geom_point(alpha = 0.7, size = 2) +
  labs(
```

```

title = "Preço vs Potência com Tipo de Transmissão",
x = "Potência (cv)",
y = "Preço (R$)",
color = "Transmissão"
) +
theme_minimal()

```

