

Análise de Eficiência de Combustível em Veículos

Alison Cordeiro Sousa

Introdução

Análise de Eficiência de Combustível em Veículos

Este estudo investiga os fatores que influenciam o consumo de combustível (mpg) em uma amostra de 26 veículos, utilizando peso do carro (weight1), preço e origem (nacional/importado) como preditores. O modelo simples peso vs consumo mostrou forte relação negativa ($\beta = -5.53$, $p < 0.001$), indicando que cada aumento de 1 tonelada reduz em 5.5 milhas por galão, explicando 65.3% da variação. A análise gráfica revelou padrão linear claro, porém com alguns outliers que merecem investigação. O modelo múltiplo, apesar de incluir preço e origem, não melhorou significativamente o ajuste (R^2 ajustado = 63.1% vs 63.9% do modelo simples, $p = 0.485$ na ANOVA), sugerindo que o peso é o principal determinante do consumo.

Implicações e Limitações

Os resultados destacam o peso como fator crítico para eficiência energética em veículos, com efeito robusto mesmo após controle por outras variáveis. A não significância do preço ($p = 0.404$) e origem ($p = 0.236$) no modelo múltiplo sugere que estas características têm impacto marginal no consumo quando considerado o peso. Os diagnósticos mostraram resíduos com distribuição aproximadamente normal, porém o pequeno tamanho amostral ($n=26$) limita a generalização dos achados. Estudos futuros deveriam incluir variáveis como potência do motor e tipo de transmissão para melhor compreensão dos determinantes da eficiência veicular.

```
# -----  
# Configuração inicial: definindo diretório e carregando dados  
# -----  
  
setwd("C:/Users/PC GAMER/Downloads/data") # Ajuste para seu diretório  
auto <- read.csv("auto.csv") # Carrega a base de dados  
  
# Estrutura dos dados para verificar tipos e variáveis  
str(auto)
```

```
## 'data.frame': 26 obs. of 8 variables:  
## $ make : chr "AMC" "AMC" "AMC" "Audi" ...  
## $ mpg : int 22 17 22 17 23 25 20 15 18 26 ...  
## $ weight : int 2930 3350 2640 2830 2070 2650 3250 4080 3670 2230 ...  
## $ weight1: num 2.93 3.35 2.64 2.83 2.07 2.65 3.25 4.08 3.67 2.23 ...  
## $ price : int 4099 4749 3799 9690 6295 9735 4816 7827 5788 4453 ...  
## $ foreign: int 0 0 0 1 1 1 0 0 0 0 ...
```

```
## $ repairs: int 3 3 3 5 3 4 3 4 3 3 ...
## $ length : int 186 173 168 189 174 177 196 222 218 170 ...
```

```
# -----
# Estatísticas descritivas das variáveis relevantes
# -----

summary(auto[, c("mpg", "weight1", "price", "foreign")]) # Sumário das variáveis
```

```
##      mpg      weight1      price      foreign
## Min.   :14.00   Min.    :2.020   Min.    : 3299   Min.    :0.0000
## 1st Qu.:17.25   1st Qu.:2.643   1st Qu.: 4466   1st Qu.:0.0000
## Median :21.00   Median :3.200   Median : 5146   Median :0.0000
## Mean   :20.92   Mean    :3.099   Mean    : 6652   Mean    :0.2692
## 3rd Qu.:23.00   3rd Qu.:3.610   3rd Qu.: 8054   3rd Qu.:0.7500
## Max.   :35.00   Max.    :4.330   Max.    :15906   Max.    :1.0000
```

```
# Matriz de correlação entre variáveis numéricas (excluindo variável categórica foreign)
cor(auto[, c("mpg", "weight1", "price")])
```

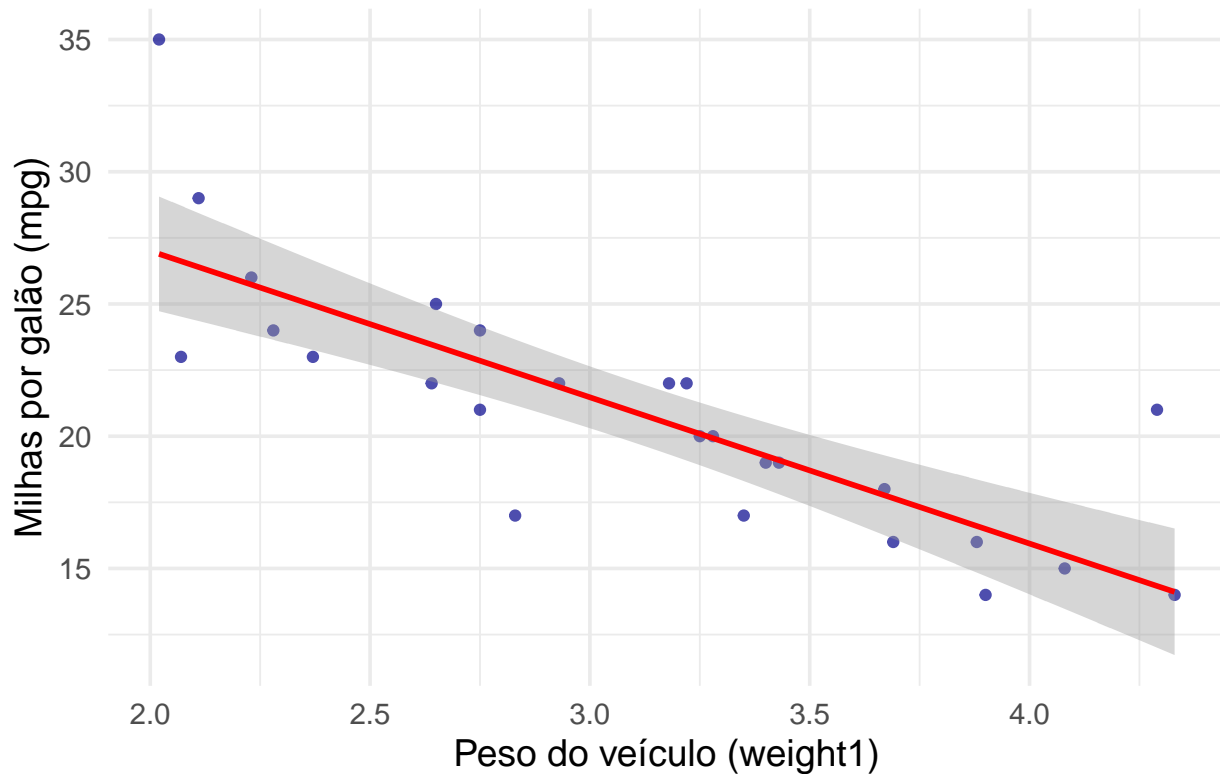
```
##      mpg      weight1      price
## mpg      1.0000000 -0.8081609 -0.4384618
## weight1 -0.8081609  1.0000000  0.5560715
## price   -0.4384618  0.5560715  1.0000000
```

```
# -----
# Gráfico de dispersão mpg x weight1 com linha de regressão
# -----

library(ggplot2)

ggplot(auto, aes(x = weight1, y = mpg)) +
  geom_point(color = "darkblue", alpha = 0.7) + # Pontos de dados
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  # Linha de regressão com intervalo de confiança
  labs(title = "Relação entre Peso do Veículo (weight1) e Consumo (mpg)",
       x = "Peso do veículo (weight1)",
       y = "Milhas por galão (mpg)") +
  theme_minimal(base_size = 14)
```

Relação entre Peso do Veículo (weight1) e Consumo (mpg)



```
# -----
# Modelo de regressão linear simples: mpg ~ weight1
# -----

modelo_simples <- lm(mpg ~ weight1, data = auto)

summary(modelo_simples) # Resumo do modelo com coeficientes, R², p-valores

##
## Call:
## lm(formula = mpg ~ weight1, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4123 -1.6073 -0.1043  0.9261  8.1072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.0665     2.6112   14.578 2.02e-13 ***
## weight1      -5.5315     0.8229   -6.722 5.93e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

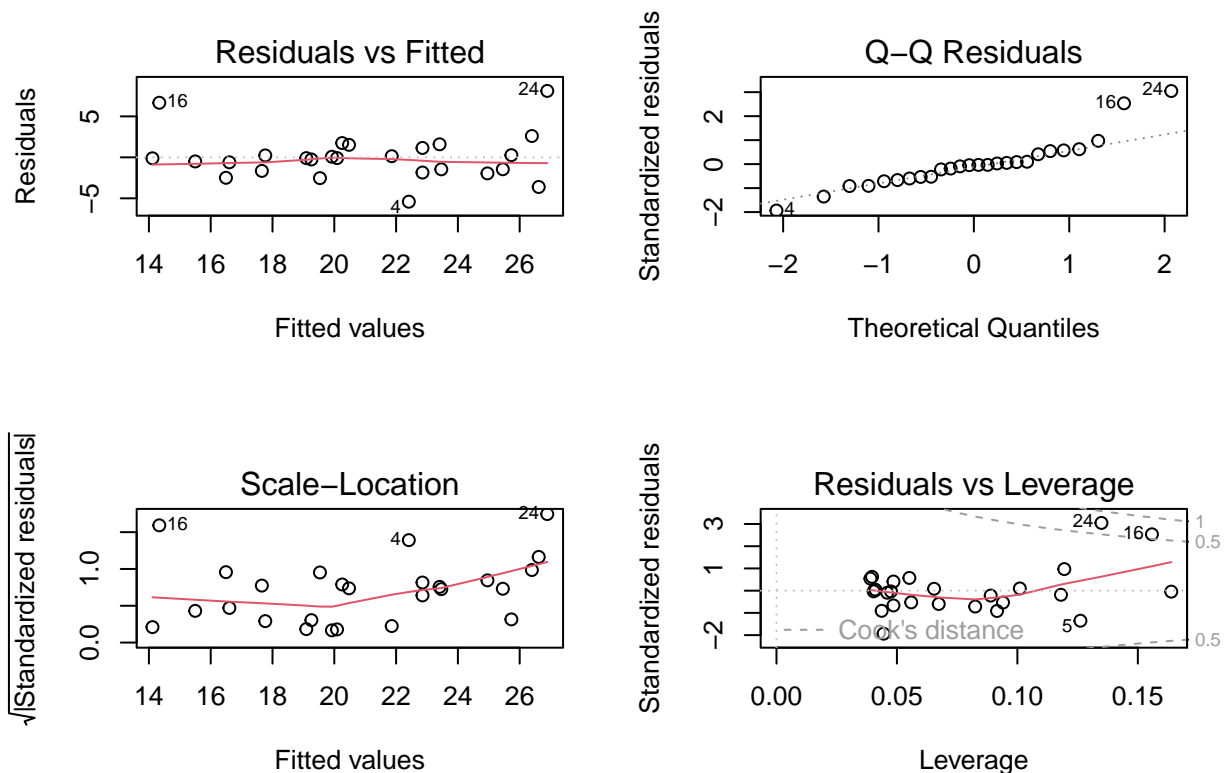
```
##
## Residual standard error: 2.86 on 24 degrees of freedom
## Multiple R-squared:  0.6531, Adjusted R-squared:  0.6387
## F-statistic: 45.19 on 1 and 24 DF,  p-value: 5.935e-07
```

```
confint(modelo_simples) # Intervalos de confiança dos coeficientes
```

```
##                2.5 %    97.5 %
## (Intercept) 32.677256 43.455664
## weight1     -7.229797 -3.833196
```

```
# -----
# Diagnóstico do modelo simples
# -----
```

```
par(mfrow = c(2,2)) # Dispor gráficos 2x2
plot(modelo_simples) # Plot de resíduos, normalidade, leverage, etc.
```



```
par(mfrow = c(1,1)) # Reset layout gráfico
```

```
# -----
```

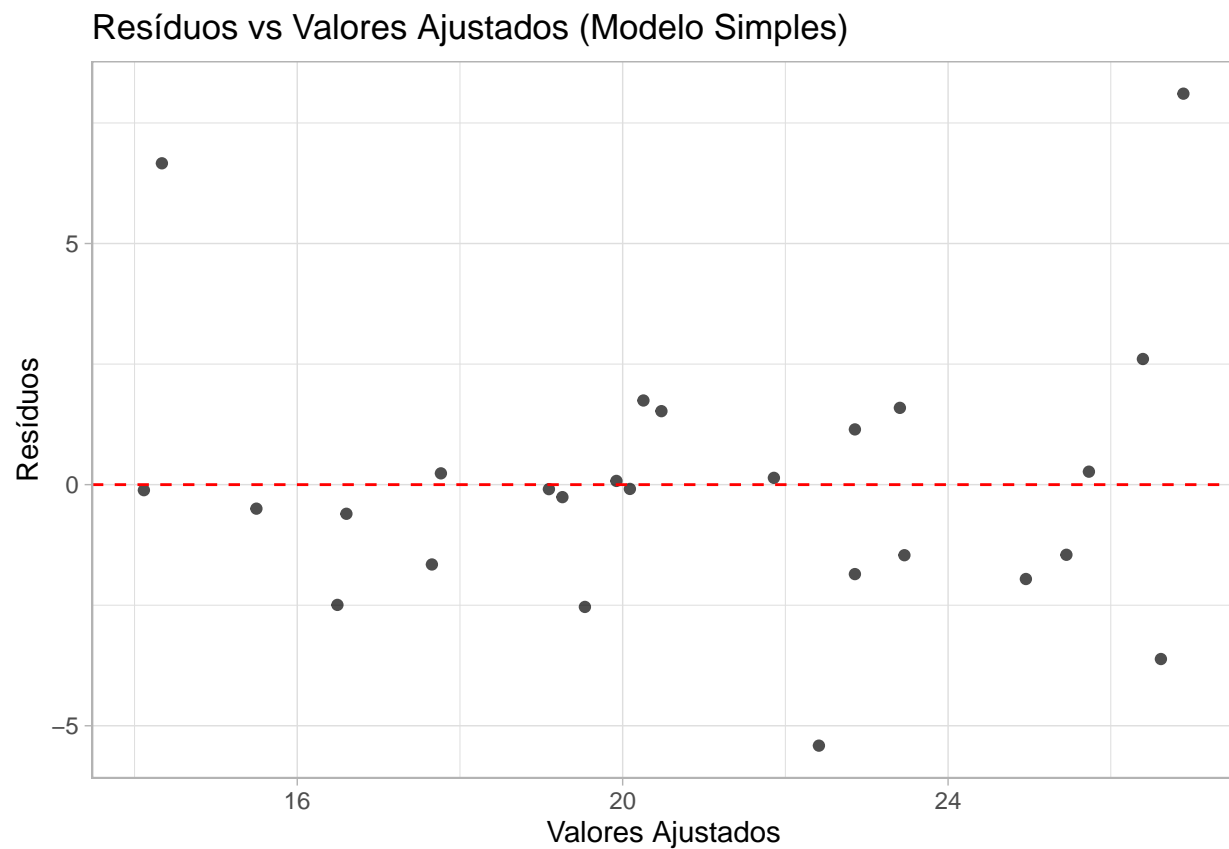
```

# Guardar valores ajustados e resíduos no dataframe para análise posterior
# -----

auto$yhat1 <- fitted(modelo_simples) # Valores ajustados (preditos)
auto$res1 <- resid(modelo_simples)   # Resíduos do modelo

# Gráfico Resíduos x Valores Ajustados para o modelo simples
ggplot(auto, aes(x = yhat1, y = res1)) +
  geom_point(alpha = 0.7) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Resíduos vs Valores Ajustados (Modelo Simples)",
       x = "Valores Ajustados",
       y = "Resíduos") +
  theme_light()

```



```

# -----
# Modelo de regressão linear múltipla: mpg ~ weight1 + price + foreign
# -----

modelo_multiplo <- lm(mpg ~ weight1 + price + foreign, data = auto)

summary(modelo_multiplo) # Resumo do modelo múltiplo

```

```
##
## Call:
## lm(formula = mpg ~ weight1 + price + foreign, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6942 -1.1857 -0.0452  0.6433  8.6895
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 42.1661962  4.2647533   9.887 1.48e-09 ***
## weight1     -7.1211114  1.6046735  -4.438 0.000207 ***
## price        0.0002258  0.0002654   0.851 0.404002
## foreign     -2.5071265  2.0565685  -1.219 0.235723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.89 on 22 degrees of freedom
## Multiple R-squared:  0.6752, Adjusted R-squared:  0.6309
## F-statistic: 15.25 on 3 and 22 DF,  p-value: 1.374e-05
```

```
confint(modelo_multiplo) # Intervalos de confiança para os coeficientes
```

```
##              2.5 %       97.5 %
## (Intercept) 3.332164e+01 51.0107531780
## weight1     -1.044900e+01 -3.7932221856
## price       -3.245229e-04  0.0007760878
## foreign     -6.772188e+00  1.7579354345
```

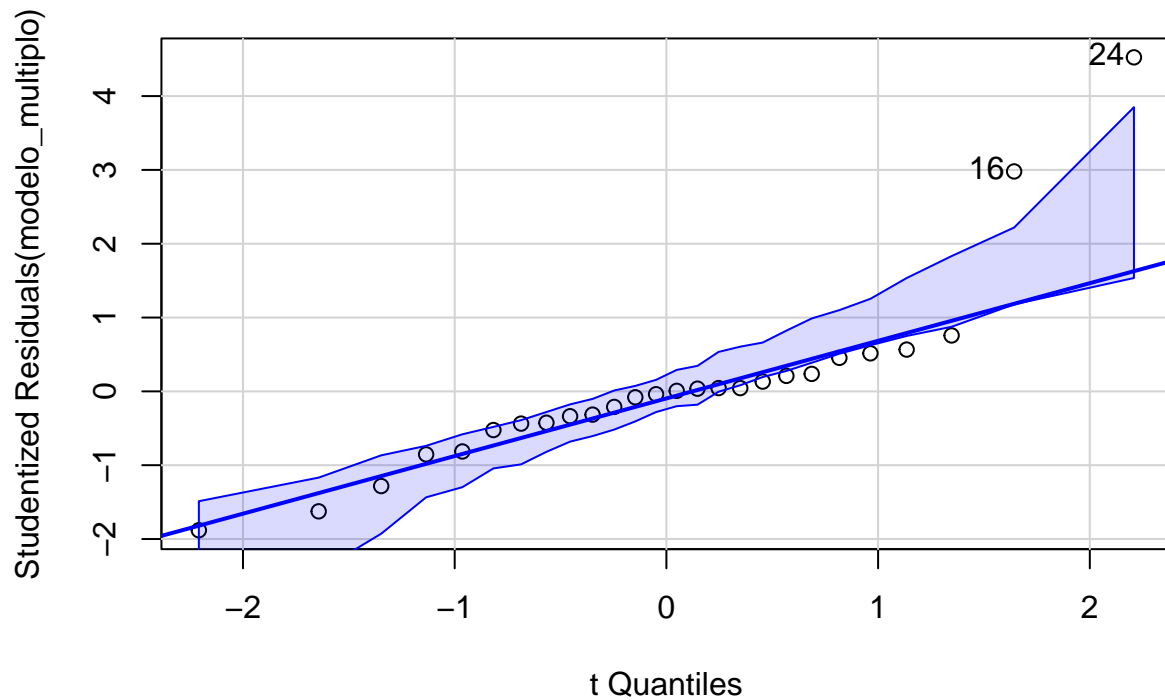
```
# -----
# Diagnóstico do modelo múltiplo
# -----

auto$yhat2 <- fitted(modelo_multiplo) # Valores ajustados do modelo múltiplo
auto$res2 <- resid(modelo_multiplo)   # Resíduos do modelo múltiplo

library(car) # Para o qqPlot

qqPlot(modelo_multiplo, main = "QQ-Plot dos Resíduos (Modelo Múltiplo)")
```

QQ-Plot dos Resíduos (Modelo Múltiplo)

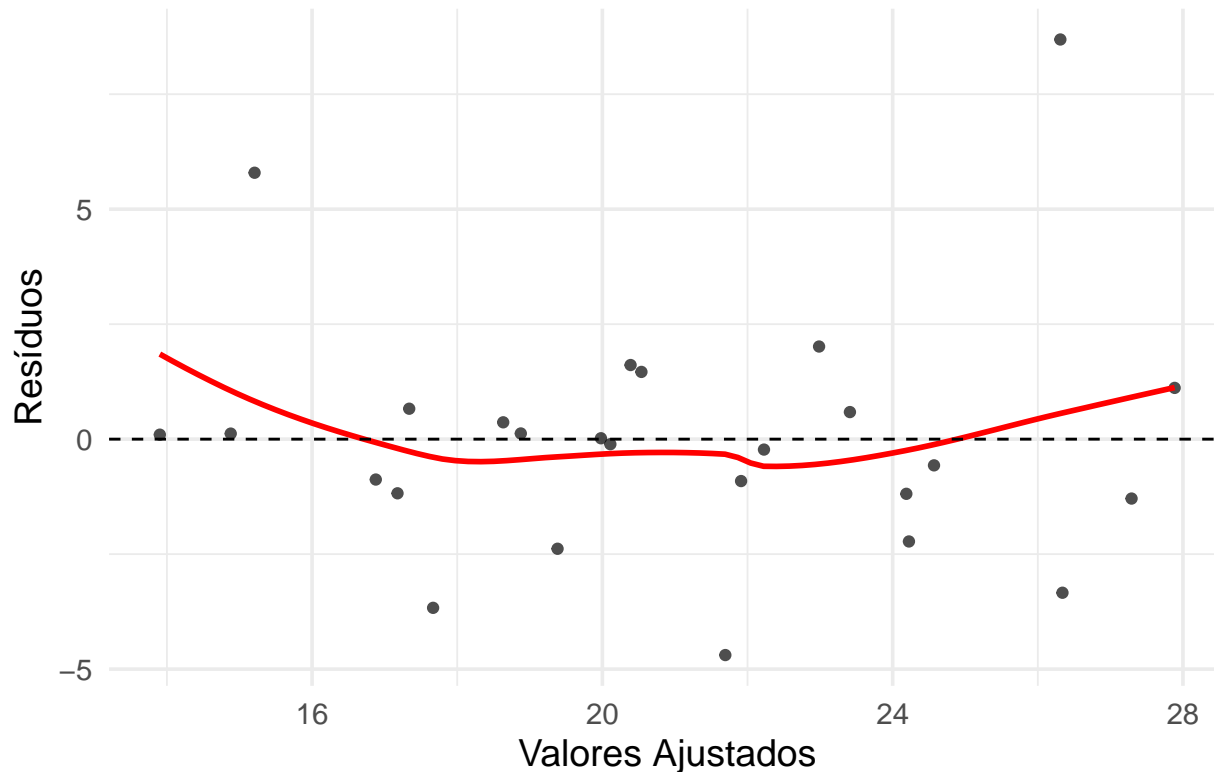


```
## [1] 16 24
```

```
# Verificar normalidade dos resíduos

# Gráfico resíduos vs valores ajustados para o modelo múltiplo
ggplot(auto, aes(x = yhat2, y = res2)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "loess", color = "red", se = FALSE) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "black") +
  labs(title = "Resíduos vs Valores Ajustados (Modelo Múltiplo)",
       x = "Valores Ajustados",
       y = "Resíduos") +
  theme_minimal(base_size = 14)
```

Resíduos vs Valores Ajustados (Modelo Múltiplo)



```
# -----
# Comparação formal dos modelos simples e múltiplo via ANOVA
# -----
```

```
anova(modelo_simples, modelo_multiplo)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ weight1
## Model 2: mpg ~ weight1 + price + foreign
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      24 196.28
## 2      22 183.77  2    12.512 0.7489 0.4845
```

```
# -----
# Comentários finais e observações:
# -----
```

```
# - O peso do veículo (weight1) é o principal determinante do consumo (mpg),
#   com coeficiente negativo e altamente significativo.
# - A inclusão das variáveis preço e origem não trouxe melhora significativa no ajuste,
#   conforme teste ANOVA ( $p > 0.4$ ).
```


- Diagnósticos gráficos indicam resíduos aproximadamente normais
ausência de violações graves dos pressupostos.
- Tamanho amostral pequeno ($n=26$) limita a generalização dos resultados;
estudos futuros devem considerar variáveis adicionais (ex: potência, transmissão).