

Regressão Linear Multivariada

EXERCÍCIO 1

Modelo de regressão: preço em função de cavalos de potência, comprimento, tamanho do motor e rendimento na cidade

```
##
## Call:
## lm(formula = price ~ horsepower + length + engine.size + city.mpg,
##     data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9697.0 -1745.7    24.9  1389.4 12904.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28480.00    7114.51  -4.003 8.99e-05 ***
## horsepower      52.74      16.62   3.174 0.001756 **
## length        114.58      32.30   3.548 0.000491 ***
## engine.size    115.32      12.92   8.922 4.06e-16 ***
## city.mpg       61.51      83.05   0.741 0.459849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3499 on 188 degrees of freedom
## Multiple R-squared:  0.8168, Adjusted R-squared:  0.8129
## F-statistic: 209.5 on 4 and 188 DF,  p-value: < 2.2e-16
```

O modelo estimado foi:

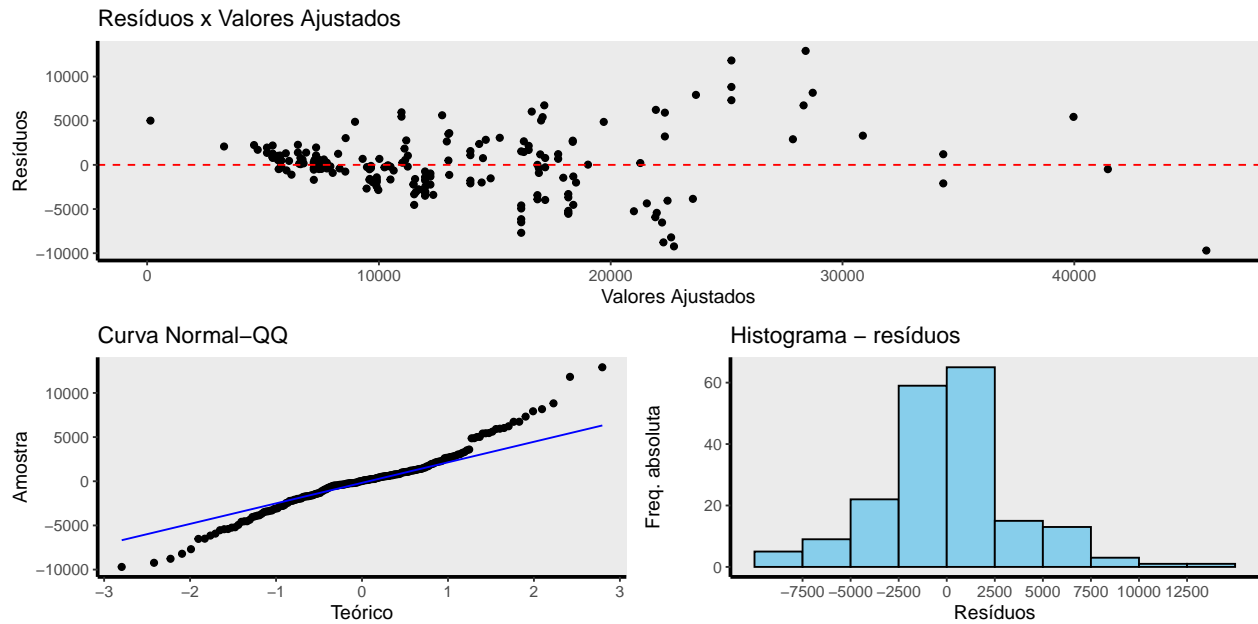
$$\hat{price}_i = -28480 + 52,74 \cdot horsepower_i + 114,58 \cdot length_i + 115,32 \cdot engine.size_i + 61,51 \cdot city.mpg_i$$

O p-valor da estatística F indica que rejeitamos a hipótese nula de ausência de significância global da regressão, isto é, pelo menos um dos coeficientes estimados é estatisticamente diferente de 0. Analisando o R^2 , vemos que 81,68% das variações no preço são explicadas pelo conjunto de variáveis explicativas utilizado.

Analisando os coeficientes individualmente, vemos que com exceção do rendimento do veículo na cidade (city.mpg), todas as variáveis são estatisticamente significativas a 1%. A análise dos coeficientes que foram significativos indica que:

1. Um aumento de uma unidade na potência está associado a um aumento de 52,74 unidades no preço;
2. Um aumento de 1 unidade no comprimento está associado a um aumento de 114,58 unidades no preço;
3. Um aumento de 1 unidade no tamanho do motor está associado a aumento de 115,32 unidades no preço.

Em seguida, temos os gráficos de diagnóstico do modelo:



Observando o gráfico dos resíduos x valores ajustados, podemos perceber que claramente ele apresenta sinais de heterocedasticidade. Realizando o teste de Breusch-Pagan, obtemos:

```
##
## studentized Breusch-Pagan test
##
## data: reg
## BP = 72.034, df = 4, p-value = 8.442e-15
```

apontando para a rejeição da hipótese nula de homocedasticidade.

Em relação a distribuição dos resíduos, enquanto as 'pontas' aparentam ser assimétricas, o centro da distribuição parece ser simétrico. Apesar disso, o conjunto dos gráficos claramente aponta para o não-atendimento da hipótese de normalidade dos resíduos.