# ANÁLISE DO CONSUMO MÉDIO DE COMBUSTÍVEL DE CARROS MODELOS 1973/1974


ELIZABETH MIE HASHIMOTO
Supervisor: Athos Damiani

Álvares Machado - Janeiro/2021

# Introdução

Na crise petrolífera de 1973, membros da Organização dos Países Árabes Exportadores de Petróleo (OPAEP) aplicaram sanções em protesto ao apoio dos Estados Unidos e outras nações à Israel durante a Guerra do Yom Kippur. O conflito resultou no aumento do preço do petróleo de três dólares por barril para cerca de 12 doláres no mundo inteiro, sendo que os preços fixados para os Estados Unidos foram ainda maiores.

Como uma alternativa à alta do preço do petróleo no mercado mundial, os Estados Unidos iniciaram um programa de eficiência energética, conhecido como *Corporate Average Fuel Economy* (CAFE), com o propósito de reduzir o consumo de combustível de carros, *pick-ups*, minivans e SUVs (Almeida Filho, 2018).

Acredita-se que a melhoria no consumo médio de combustível dos automóveis leva à redução das contas de importação de petróleo, ou seja, pode resultar em economias estimadas nas contas anuais de importação de petróleo no valor de 300 bilhões de dólares em 2025 e 600 bilhões em 2050 (Global fuel economy initiative, 2021). Por outro lado, a eficiência do combustível depende de muitas características do veículo, incluindo as especificações do motor, resistência aerodinâmica, peso, combustível e entre outros atributos.

Nesse contexto, buscou-se validar a hipótese de que modificações na estrutura do automóvel aumenta o seu consumo médio. Portanto, o presente trabalho teve como objetivo identificar quais características do carro explica a sua eficiência medida em milhas por galão. As análises foram feitas utilizando o *software* `R` versão 4.0.3, considerando um nível de significância de 5%.

# 1 Análise Exploratória

O conjunto de dados, denominado de `mtcars`, foi obtido a partir das edições de março, abril, junho e julho de 1974 da revista *Motor Trend* para um estudo realizado por Hocking (1976) e posteriormente, reportado por Henderson e Velleman (1981). Os dados, em questão, são referentes ao consumo de gasolina e dez características físicas de 32 automóveis modelos 1973-1974. O mesmo está disponível na biblioteca `datasets` do *software* `R` para consulta.

Dessa forma, de acordo com a hipótese formulada, as variáveis observadas no conjunto de dados são definidas como:

✓ **Variável resposta**

- `mpg`: eficiência (milhas por galão de combustível).

✓ **Variável explicativa**

- `cyl`: número de cilindros.
- `disp`: cilindradas (polegada cúbica).
- `hp`: potência bruta (HP).
- `drat`: relação de eixo traseiro.
- `wt`: peso (1000 libras).
- `qsec`: tempo no quarto de milha (segundos).
- `vs`: formato do motor (0 = V e 1 = linha).
- `am`: tipo de transmissão (0 = automático e 1 = manual).
- `gear`: número de marchas para frente.
- `carb`: número de carburadores.

Para reduzir as informações do conjunto de dados, estatísticas descritivas de cada uma das variáveis quantitativas foram obtidas e apresentadas na Tabela 1. Os resultados mostram que não há nenhuma dado faltante e portanto, não há necessidade de imputar valores. Além disso, em média, a eficiência dos carros é de 20,09 mpg e são carros com 2 carburadores, seis cilindros e peso de 3,22 $\times$1000 libras.

**Tabela** 1: Estatísticas descritivas das variáveis de natureza quantitativa

| Variável | Missing | Média | Desvio padrão | Mínimo | Q1 | Mediana | Q3 | Máximo |
|---|---|---|---|---|---|---|---|---|
| carb | 0 | 2.81 | 1.615 | 1.00 | 2.00 | 2.00 | 4.00 | 8.00 |
| cyl | 0 | 6.19 | 1.786 | 4.00 | 4.00 | 6.00 | 8.00 | 8.00 |
| disp | 0 | 230.72 | 123.939 | 71.10 | 120.83 | 196.30 | 326.00 | 472.00 |
| drat | 0 | 3.60 | 0.535 | 2.76 | 3.08 | 3.70 | 3.92 | 4.93 |
| gear | 0 | 3.69 | 0.738 | 3.00 | 3.00 | 4.00 | 4.00 | 5.00 |
| hp | 0 | 146.69 | 68.563 | 52.00 | 96.50 | 123.00 | 180.00 | 335.00 |
| mpg | 0 | 20.09 | 6.027 | 10.40 | 15.43 | 19.20 | 22.80 | 33.90 |
| qsec | 0 | 17.85 | 1.787 | 14.50 | 16.89 | 17.71 | 18.90 | 22.90 |
| wt | 0 | 3.22 | 0.978 | 1.51 | 2.58 | 3.33 | 3.61 | 5.42 |

Na Figura 1 é apresentada um correlograma das variáveis explicativas. Por meio do gráfico, observou-se que as variáveis explicativas apresentavam uma alta correlação, ou seja, há indicativos de problema de multicolinearidade.

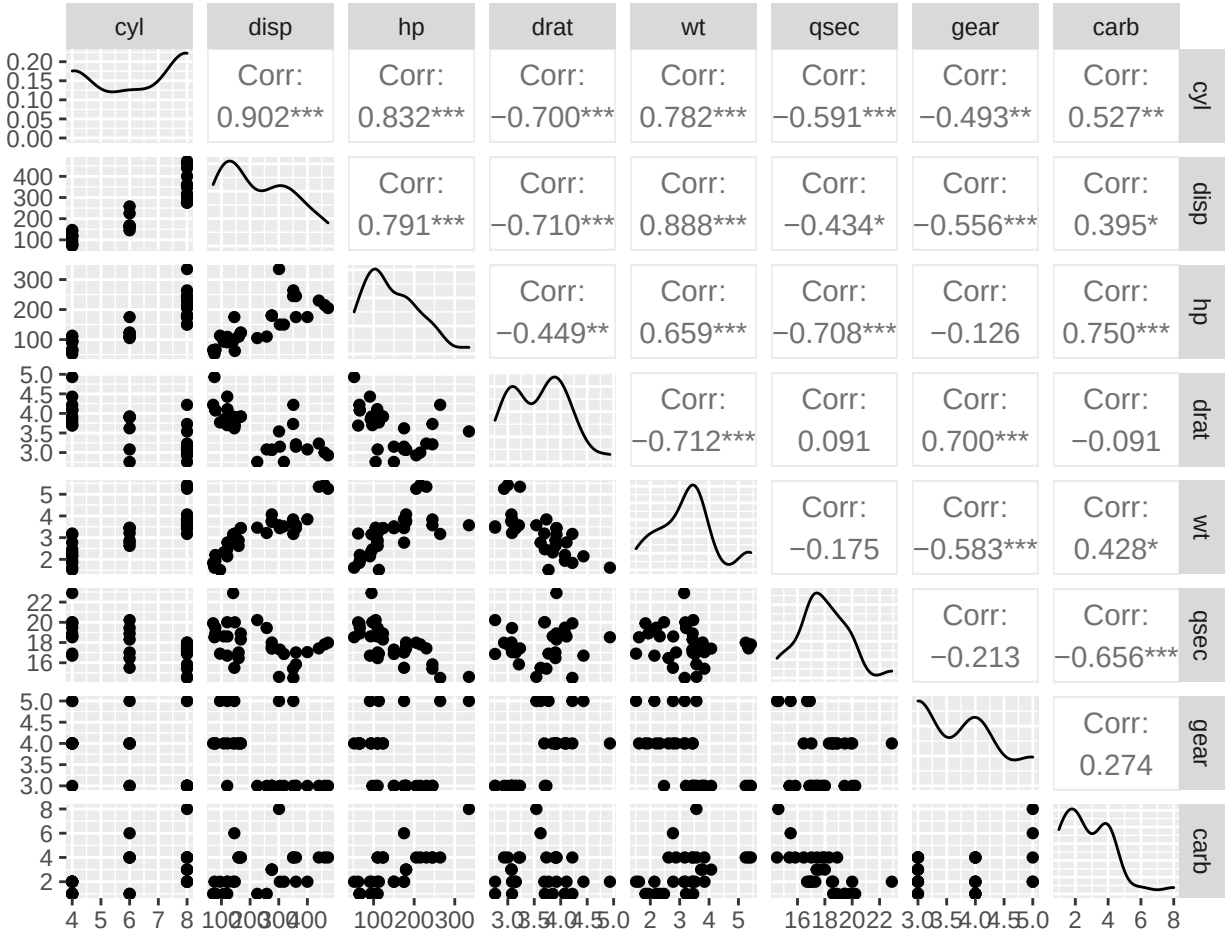


**Figura** 1: Matriz de correlação das variáveis explicativas quantitativas

Na Figura 2 são apresentados os gráficos de dispersão, na qual observou-se que uma relação linear da variável resposta com as variáveis `cyl`, `drat`, `wt` e `qsec`. Nas demais variáveis, a relação tende a ser não linear.

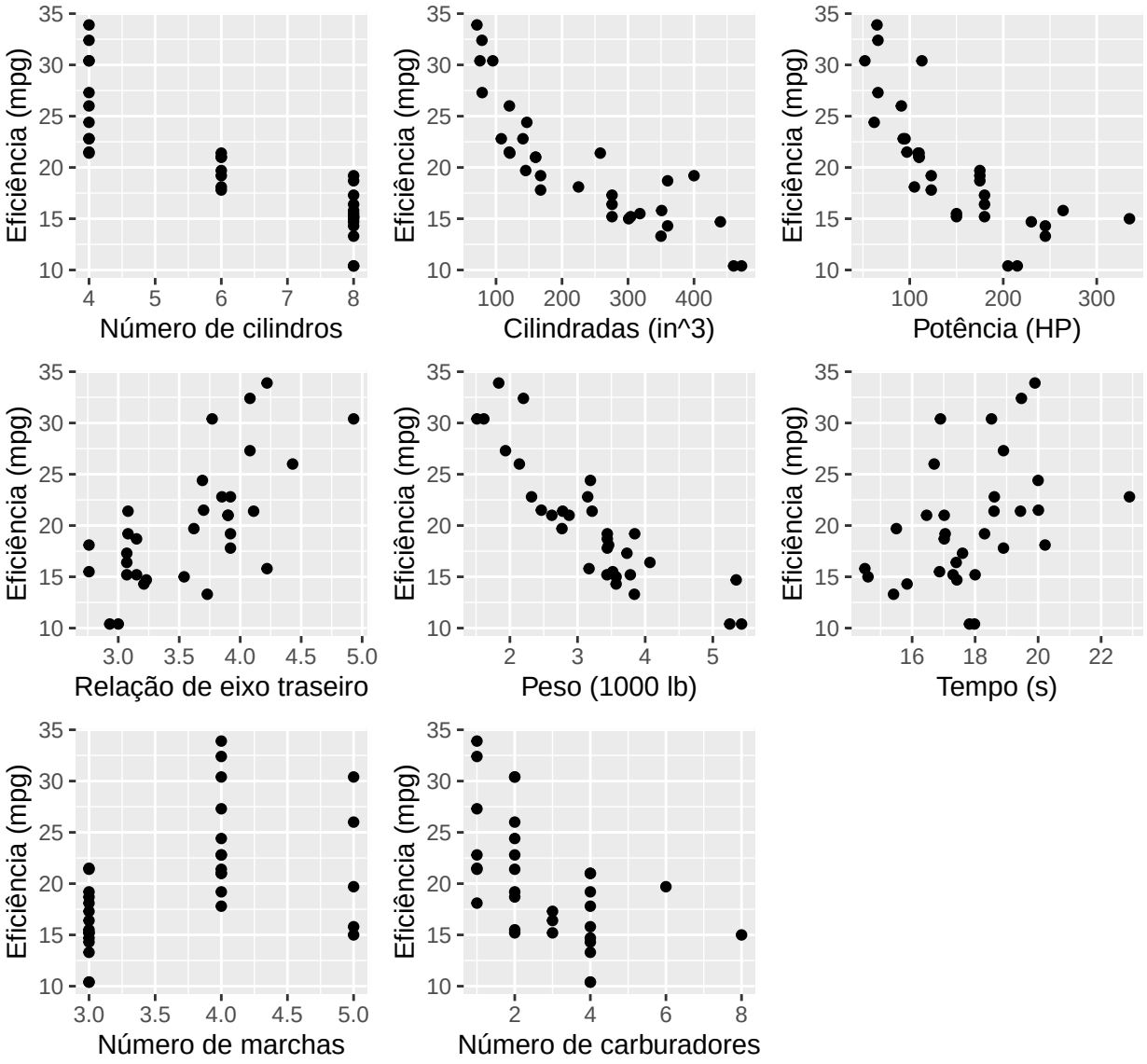


**Figura** 2: Gráfico de dispersão

Nas Figuras 3-5 são apresentados os gráficos de dispersão em função de outras covariáveis.

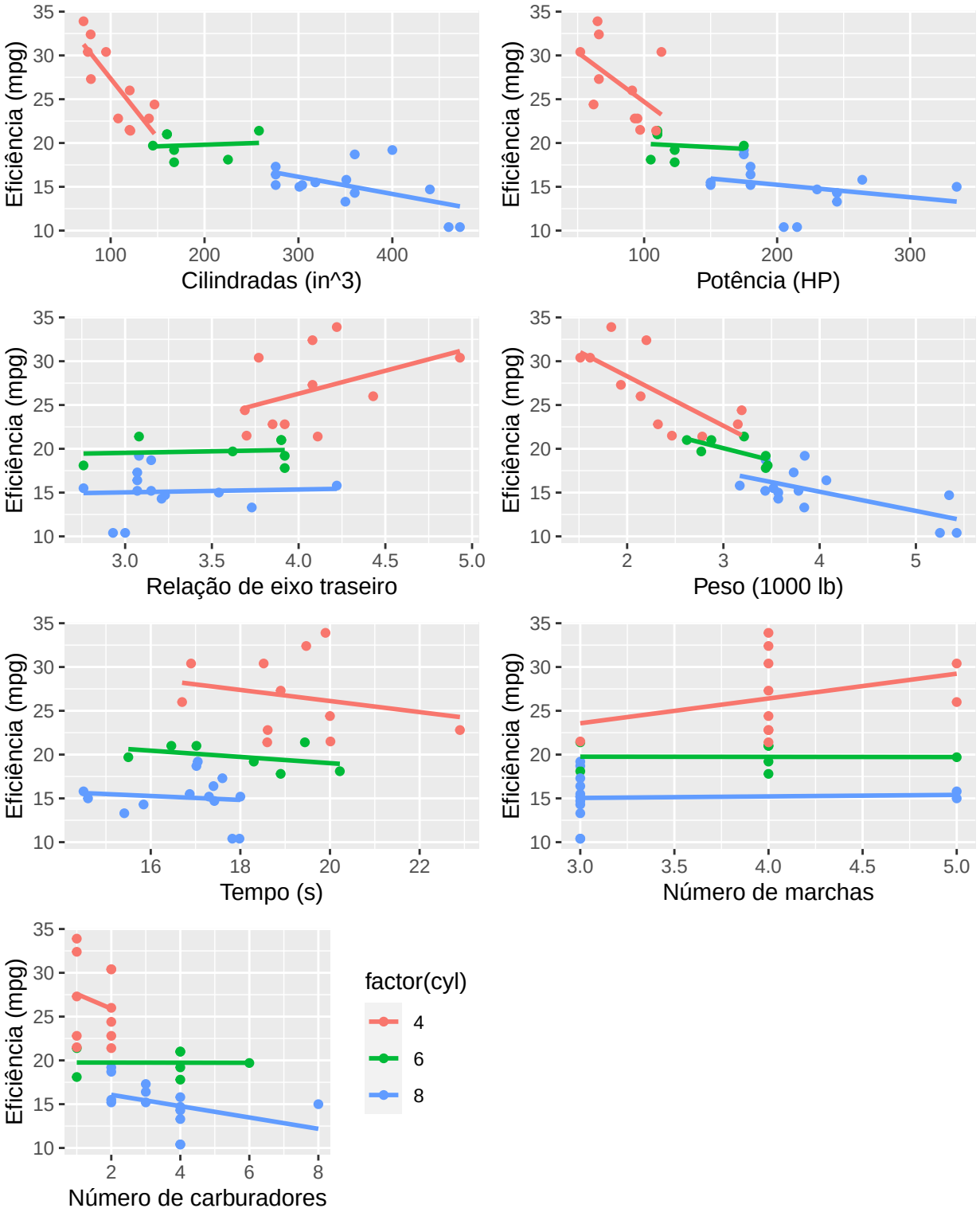


**Figura** 3: Gráfico de dispersão com pontos estratificados pelo número de cilindros

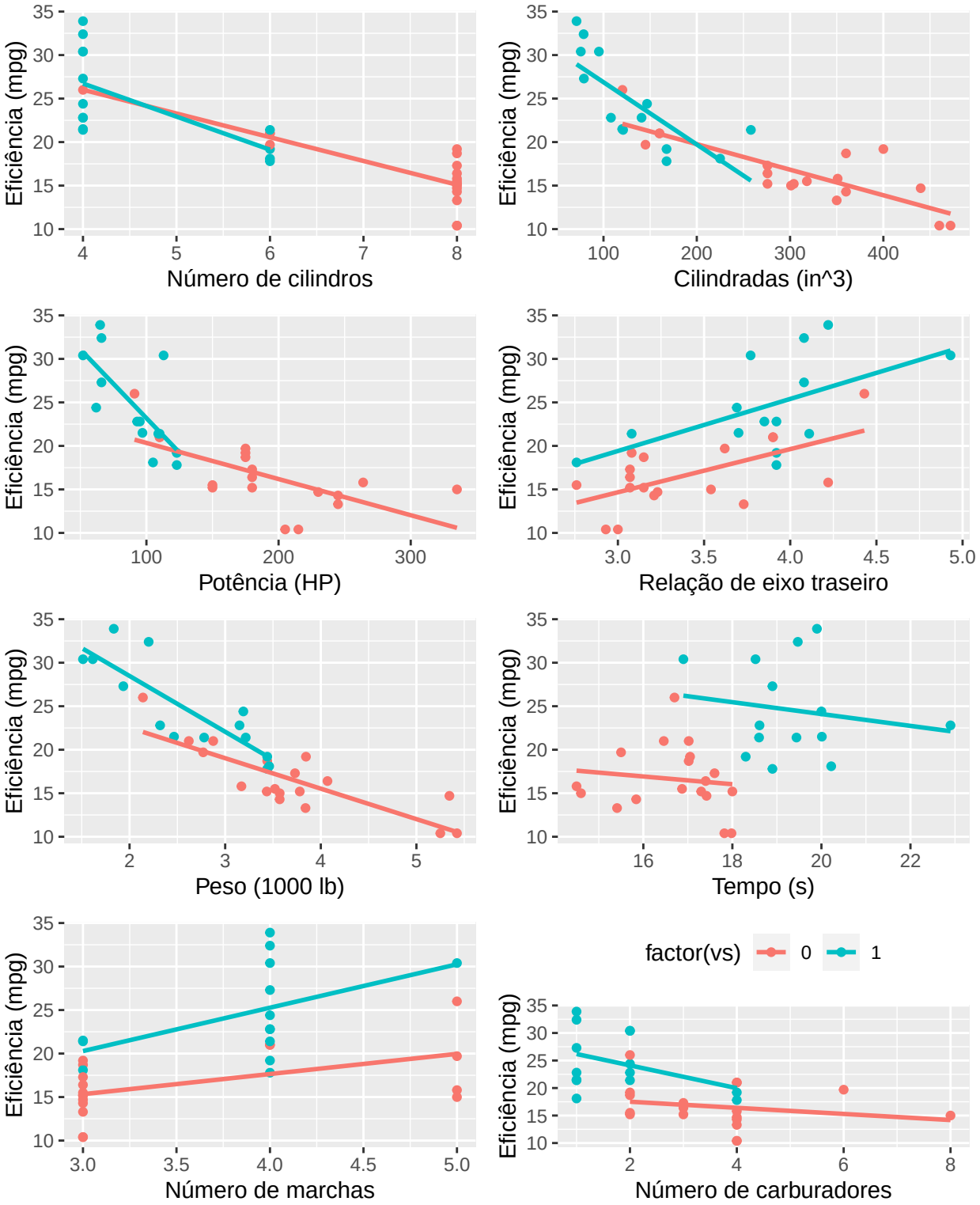


**Figura** 4: Gráfico de dispersão com pontos estratificados pelo formato do motor

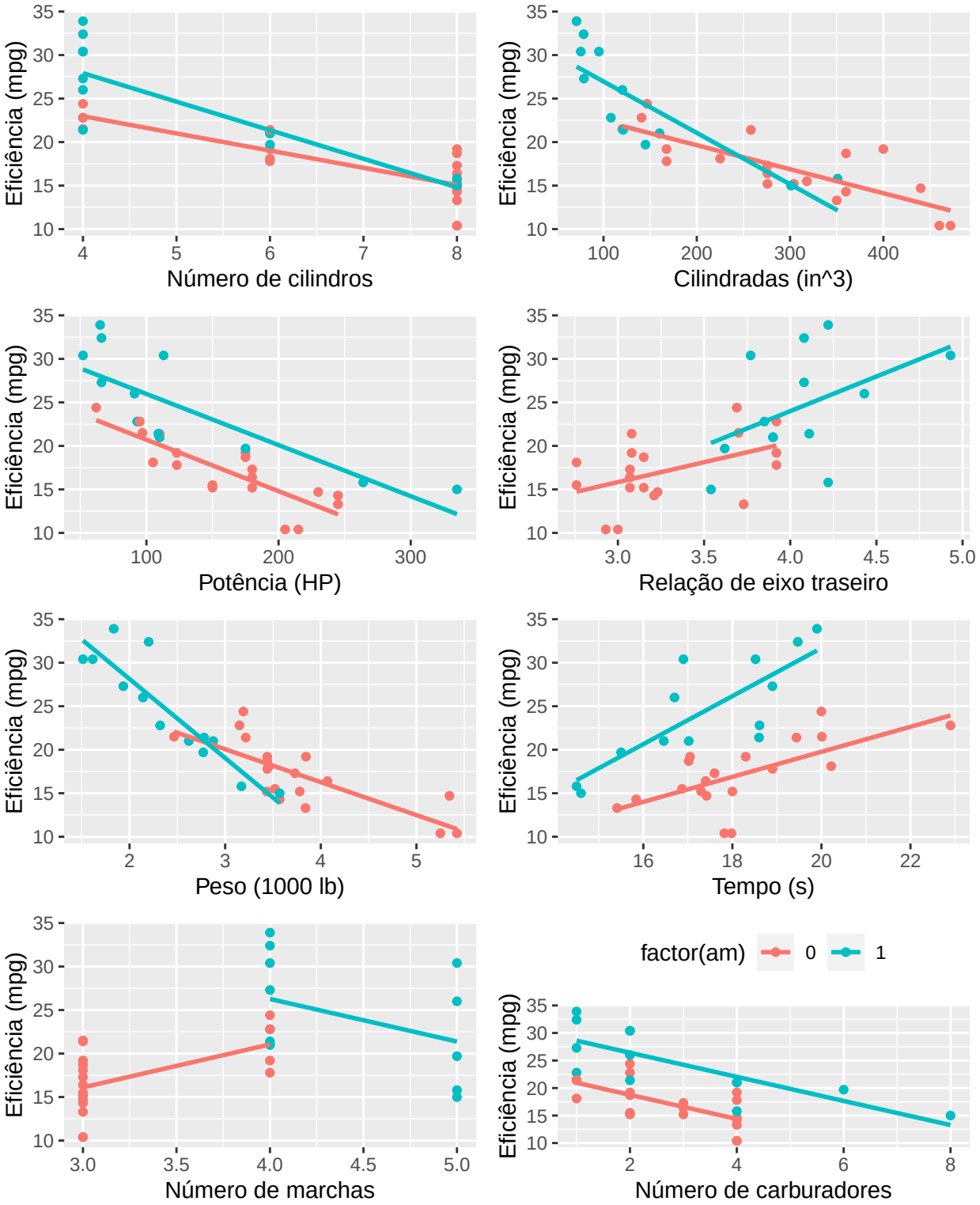


**Figura** 5: Gráfico de dispersão com pontos estratificados pelo tipo de transmissão

Na Figura 6 são apresentados os boxplots, na qual observou-se que há uma possível diferença entre o formato do motor em relação a eficiência do carro, assim como há uma diferença entre o tipo de transmissão.

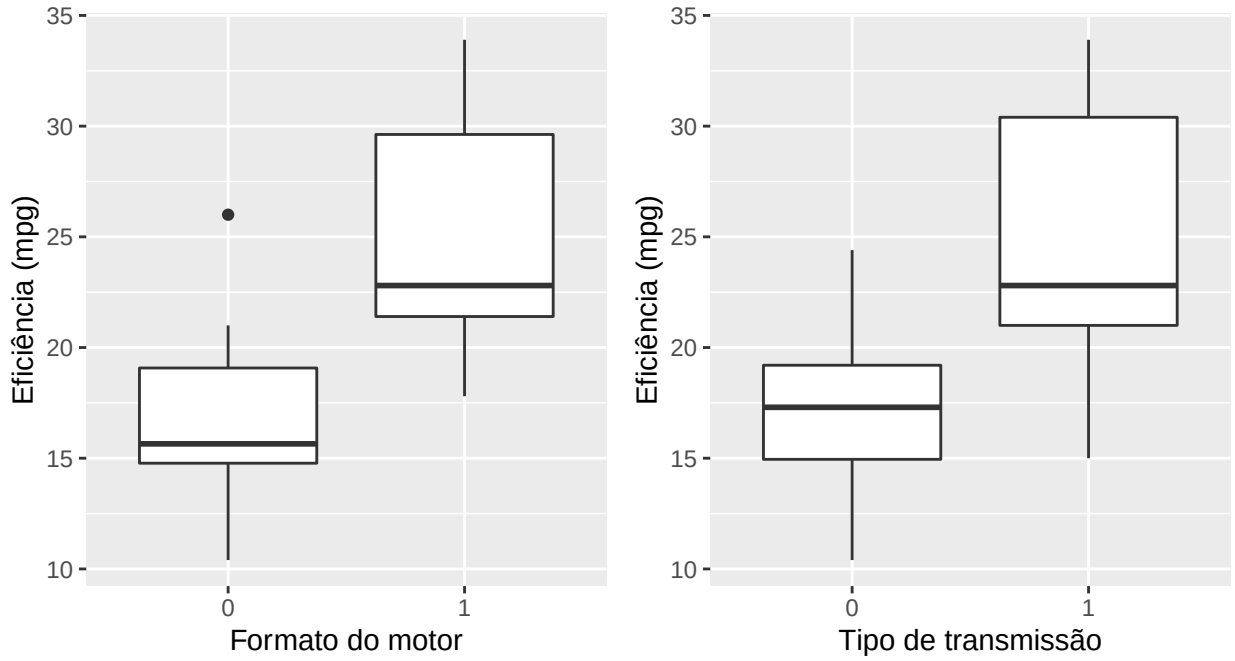


**Figura** 6: Boxplot

## 2 Modelagem

O modelo de regressão linear múltiplo, como definido em James et al. (2013) é dado por

$$\mathbf{Y} = \beta_{\mathbf{0}} + \beta_{\mathbf{1}}\mathbf{X_1} + \ldots + \beta_{\mathbf{p}}\mathbf{X_p} + \varepsilon, \tag{1}$$

em que $\mathbf{Y}$ representa a variável resposta, $\mathbf{X_1}, \ldots, \mathbf{X_p}$ é o vetor de variáveis explicativas, $\beta_0, \ldots, \beta_p$ são os parâmetros a serem estimados e $\varepsilon$ é o vetor de termos aleatórios do modelo.

Então, dado o conjunto de dados `mtcars`, o modelo de regressão (1), reescrito em função do conjunto de dados `mtcars` é dado por

$$mpg_i = \beta_0 + \beta_1 cyl_i + \beta_2 disp_i + \beta_3 hp_i + \beta_4 drat_i + \beta_5 wt_i + \ldots + \beta_9 gear_i + \beta_{10} carb_i + \varepsilon_i, \quad i = 1, \ldots, 32,$$

sendo este, denominado de modelo completo. Os parâmetros $\beta_0, \ldots, \beta_{10}$ foram estimados pelo método de mínimos quadrados com auxílio computacional do *software* `R`. Além disso, o efeito das variáveis explicativadas sobre a variável resposta {mpg} foram testadas considerando as seguintes hipóteses estatísticas

$$H_0 : \beta_j = 0 \quad vs \quad H_a : \beta_j \neq 0, j = 0, 1, \ldots, 10.$$

Nesse cenário, obteve-se os seguintes resultados:

✓ **Modelo completo**

As estimativas, bem os erros padrões e os $p - valores$ dos parâmetros do modelo completo foram obtidas pelo código

```
mod_completo <- lm(mpg ~ ., data=mtcars)
summary(mod_completo)
```

e apresentadas na Tabela 2. Os resultados apresentados nessa tabela indicam que nenhuma da variáveis explicativas tem alguma relação com a eficiência do carro, pois o $p$-valor é maior do o nível de significância e consequentemente, levando a rejeição da hipótese nula.

Por outro lado, na análise exploratório foi identificado o problema de multicolinearidade. Dessa forma, o fator de inflação da variância (VIF) foi calculo pelo seguinte código

```
car::vif(mod_completo)
```

e os valores apresentados na Tabela 3. Segundo James et al. (2013), variáveis explicativas cujo VIF for maior do que cinco podem ser removidas do modelo como uma das soluções para o problema. Nessa situação, de acordo com a Tabela 3, as variáveis explicativas `drat`, `vs`, `am` e `wt` foram mantidas no modelo. Justifica-se a permanência da variável `wt` em função do comportamento linear quando comparado com as demains variáveis quantativas contínuas (Figura 2).

**Tabela** 2: Estimativas dos parâmetros do modelo completo

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 12.303 | 18.718 | 0.657 | 0.518 |
| cyl | -0.111 | 1.045 | -0.107 | 0.916 |
| disp | 0.013 | 0.018 | 0.747 | 0.463 |
| hp | -0.021 | 0.022 | -0.987 | 0.335 |
| drat | 0.787 | 1.635 | 0.481 | 0.635 |
| wt | -3.715 | 1.894 | -1.961 | 0.063 |
| qsec | 0.821 | 0.731 | 1.123 | 0.274 |
| vs | 0.318 | 2.105 | 0.151 | 0.881 |
| am | 2.520 | 2.057 | 1.225 | 0.234 |
| gear | 0.655 | 1.493 | 0.439 | 0.665 |
| carb | -0.199 | 0.829 | -0.241 | 0.812 |

**Tabela** 3: Fator de inflação da variância das variáveis explicativas

| | VIF |
|---|---|
| cyl | 15.37 |
| disp | 21.62 |
| hp | 9.83 |
| drat | 3.38 |
| wt | 15.16 |
| qsec | 7.53 |
| vs | 4.97 |
| am | 4.65 |
| gear | 5.36 |
| carb | 7.91 |

### ✓ Modelo reduzido

- `mod_red0`: modelo reduzido 1

$$mpg_i = \beta_0 + \beta_4 drat_i + \beta_5 wt_i + \beta_7 vs_i + \beta_8 am_i + \varepsilon_i, \quad i = 1, \ldots, 32,$$

cujas estimativas dos parâmetros são obtidas por meio do seguinte código

```
mod_red1 <- lm(mpg ~ drat + wt + vs + am, data=mtcars)
summary(mod_red1)
```

e os resultados apresentados na Tabela 4. Como a hipótese nula não foi rejeitada para os coeficientes associados as variáveis `wt` ($p$-valor = 0,000) e `vs` ($p$-valor = 0,016), ou seja, o peso e o formato do motor tem um possível efeito sobre a eficiência do carro. Entretanto, especialistas em mecânica acreditam o tipo de transmissão tem alguma influência sobre o consumo médio de um automável. Por essa razão, a variável `am` ainda foi mantida no modelo. Além disso, na Figura 5 foi observado uma possível interação entre o tipo de transmissão e o peso do carro.

**Tabela** 4: Estimativas dos parâmetros do modelo reduzido 1

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 27.573 | 6.874 | 4.011 | 0.000 |
| drat | 0.682 | 1.559 | 0.438 | 0.665 |
| wt | -3.699 | 0.932 | -3.967 | 0.000 |
| vs | 3.452 | 1.348 | 2.561 | 0.016 |
| am | 1.115 | 1.736 | 0.642 | 0.526 |

- `mod_red2`: modelo reduzido 2

$$mpg_i = \beta_0 + \beta_7 vs_i + \beta_5 wt_i + \beta_8 am_i + \beta_{58} wt_i \times am_i + \varepsilon_i, \quad i = 1, \ldots, 32,$$

cujas estimativas dos parâmetros são obtidas por meio do seguinte código

```
mod_red2 <- lm(mpg ~ vs + wt*am, data=mtcars)
summary(mod_red2)
```

e os resultados apresentados na Tabela 5. De acordo com essa tabela, a hipótese nula foi rejeitada em todos os casos, pois o $p$-valor foi menor do que o nível de significância. Dessa forma, o formato do motor e assim como a interação entre peso e tipo de transmissão tem algum efeito sobre a eficiência do carro.

**Tabela** 5: Estimativas dos parâmetros do modelo reduzido 2

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 26.25 | 3.346 | 7.85 | 0.000 |
| vs | 2.93 | 1.095 | 2.68 | 0.012 |
| wt | -2.70 | 0.818 | -3.30 | 0.003 |
| am | 14.32 | 3.866 | 3.70 | 0.001 |
| wt:am | -4.66 | 1.329 | -3.51 | 0.002 |

- `mod_red3`: modelo reduzido 3

Levando em consideração a análise exploratória e a opinião de especialistas em mecânica, um modelo alternativo é dado por

$$mpg_i = \beta_0 + \beta_5 wt_i + \beta_{16} cyl6_i + \beta_{18} cyl8_i + \beta_{56} wt_i \times cyl6_i + \beta_{58} wt_i \times cyl8_i + \varepsilon_i, \quad i = 1, \ldots, 32,$$

cujas estimativas dos parâmetros são obtidas por meio do seguinte código

```
mod_red3 <- lm(mpg ~ wt*factor(cyl), data=mtcars)
summary(mod_red3)
```

e os resultados apresentados na Tabela 6. Como a variável `cyl` foi categorizada para simplificar a interpertação do efeito da interação, duas variáveis *dummies* foram criadas, assumindo a categoria `cyl4` como casela de referência. Logo, verificou-se uma possível relação do peso, assim como do efeito da interação entre peso e número de cilindro sobre a eficiência do carro, uma vez que a hipótese de nula foi rejeitada.

**Tabela** 6: Estimativas dos parâmetros do modelo reduzido 3

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 39.57 | 3.19 | 12.39 | 0.000 |
| wt | -5.65 | 1.36 | -4.15 | 0.000 |
| factor(cyl)6 | -11.16 | 9.36 | -1.19 | 0.244 |
| factor(cyl)8 | -15.70 | 4.84 | -3.25 | 0.003 |
| wt:factor(cyl)6 | 2.87 | 3.12 | 0.92 | 0.366 |
| wt:factor(cyl)8 | 3.46 | 1.63 | 2.12 | 0.043 |

Na Tabela 7 são apresentados os valores de R2 e R2 ajustado para os três modelos reduzidos estimados. Os valores obtidos indicam que o modelo `mod_red2` é mais adequado entre os três modelos estimados, seguido do modelo `mod_red3` e `mod_red1`. Entretanto, os valores de R2 ajustados não são suficientes para determinar a adequação do modelo. À vistsa disso, uma análise de resíduo foi realizada.

**Tabela** 7: Valores de R2 e R2 ajustados dos modelos reduzidos

| Modelo | R2 | R2_ajustado |
|---|---|---|
| mod_red1 | 0.809 | 0.781 |
| mod_red2 | 0.868 | 0.849 |
| mod_red3 | 0.862 | 0.835 |

# 3 Diagnóstico do Modelo

## 3.1 Dados completo

Para cada modelo reduzido estimado foi realizado uma análise de resíduo e os resultados são apresentados nas Figuras 7, 8 e 9, respectivamente.

✓ **Modelo reduzido 1**
De acordo com a Figura 7:

- **Residuals vs Fitted**: uma leve semelhança com uma parábola com concavidade voltada para cima, ou seja, temos um possível padrão não linear entre as variáveis.
- **Normal Q-Q**: A maior parte dos pontos encontra-se em torno da linha tracejada, exceto pelos pontos `Chrysler Imperial`, `Toyota Corolla`, `Fiat 128` e outros dois pontos não identificados na parte inferior da figura. O que indica que esses três pontos são possíveis *outliers*.
- **Scale-Location**: aparentemente os resíduos aparecem espalhados aleatoriamente, o que indica que a suposição de homocedasticidade é satisfeita, ou seja, a variância é constante.
- **Residuals vs Leverage**: como todos os pontos são menores do que a distância de Cook, temos evidências de que não há pontos de alavanca.

✓ **Modelo reduzido 2**
De acordo com a Figura 8:

- **Residuals vs Fitted**: os resíduos não mostram nenhum padrão, uma vez que a linha vermelha se parece com uma linha reta.
- **Normal Q-Q**: A maior parte dos pontos encontra-se afastada da linha tracejada e além disso, os carros `Merc 240D`, `Toyota Corolla` e `Fiat 128` foram apontados como possíveis *outliers*.
- **Scale-Location**: aparentemente os resíduos aparecem espalhados aleatoriamente, o que indica que a suposição de homocedasticidade é satisfeita, ou seja, a variância é constante.

- **Residuals vs Leverage**: como todos os pontos são menores do que a distância de Cook, temos evidências de que não há pontos de alavanca.

✓ **Modelo reduzido 3**

De acordo com a Figura 9:

- **Residuals vs Fitted**: os resíduos não mostram nenhum padrão, uma vez que a linha vermelha se parece com uma linha reta.
- **Normal Q-Q**: Alguns pontos encontra-se afastada da linha tracejada e além disso, os carros `Toyota Corona`, `Toyota Corolla` e `Fiat 128` foram apontados como possíveis *outliers*.
- **Scale-Location**: aparentemente os resíduos aparecem espalhados aleatoriamente, o que indica que a suposição de homocedasticidade é satisfeita, ou seja, a variância é constante.
- **Residuals vs Leverage**: como todos os pontos são menores do que a distância de Cook, temos evidências de que não há pontos de alavanca.

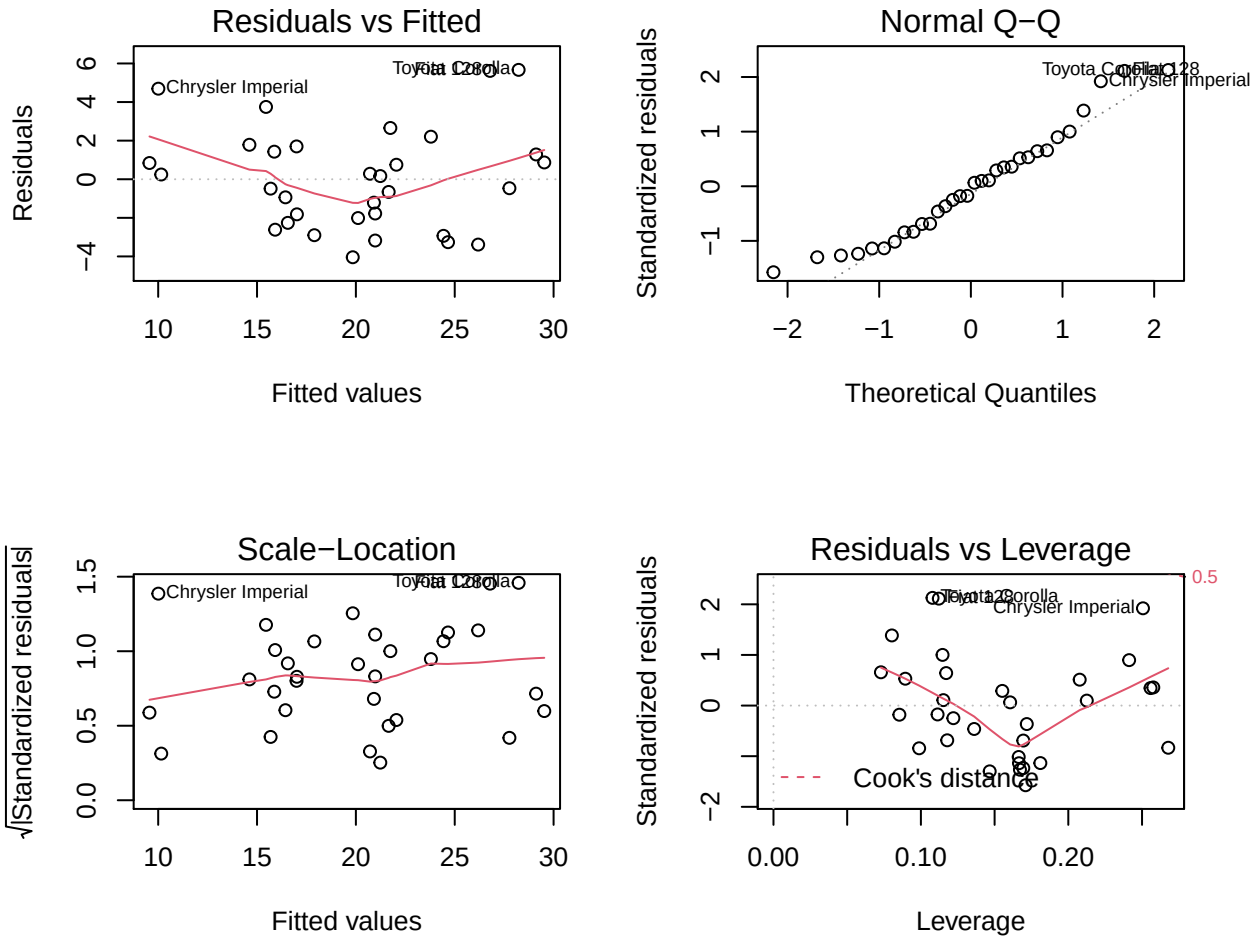


**Figura** 7: Gráfico de resíduos do modelo reduzido 1

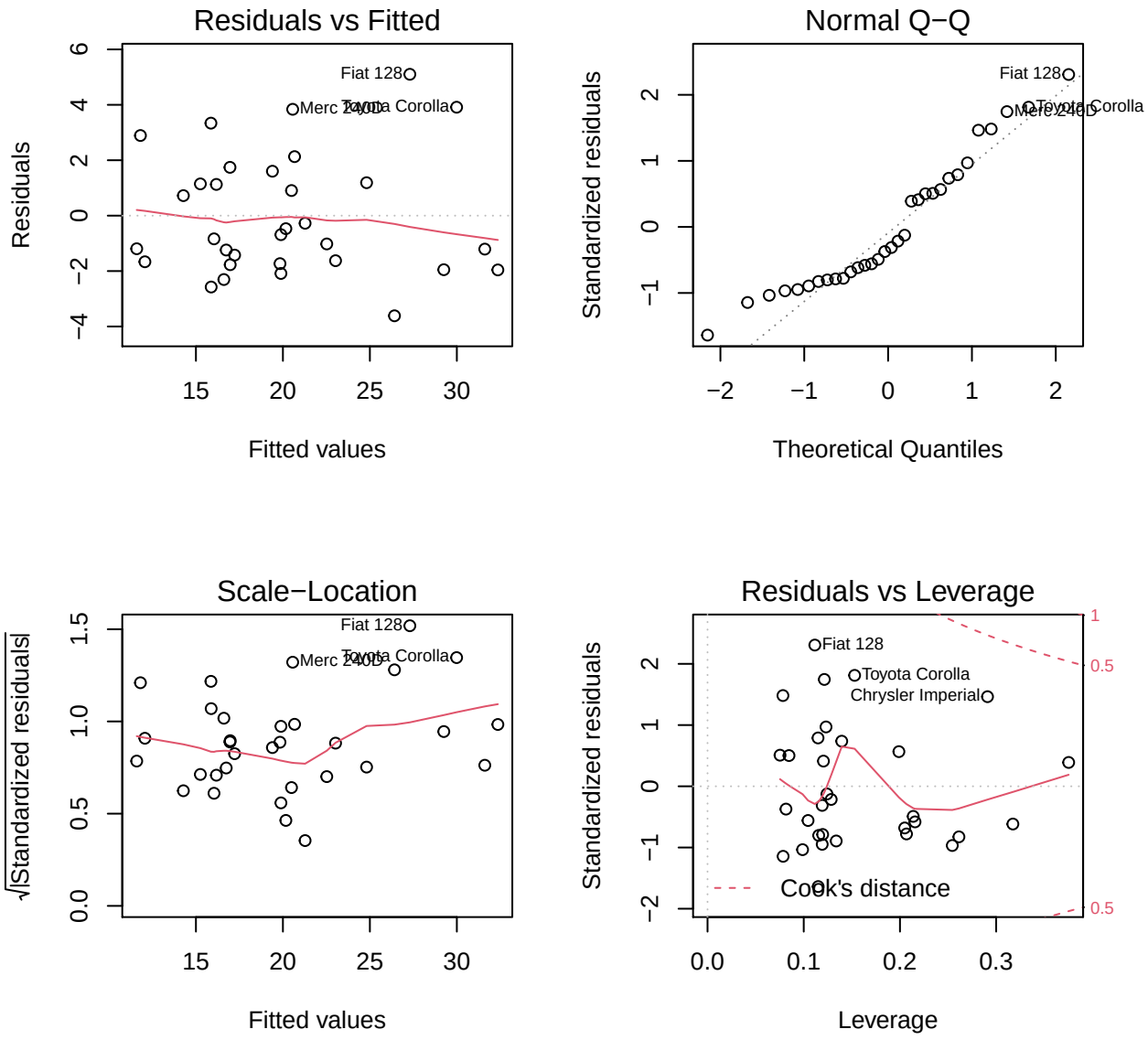


**Figura** 8: Gráfico de resíduos do modelo reduzido 2

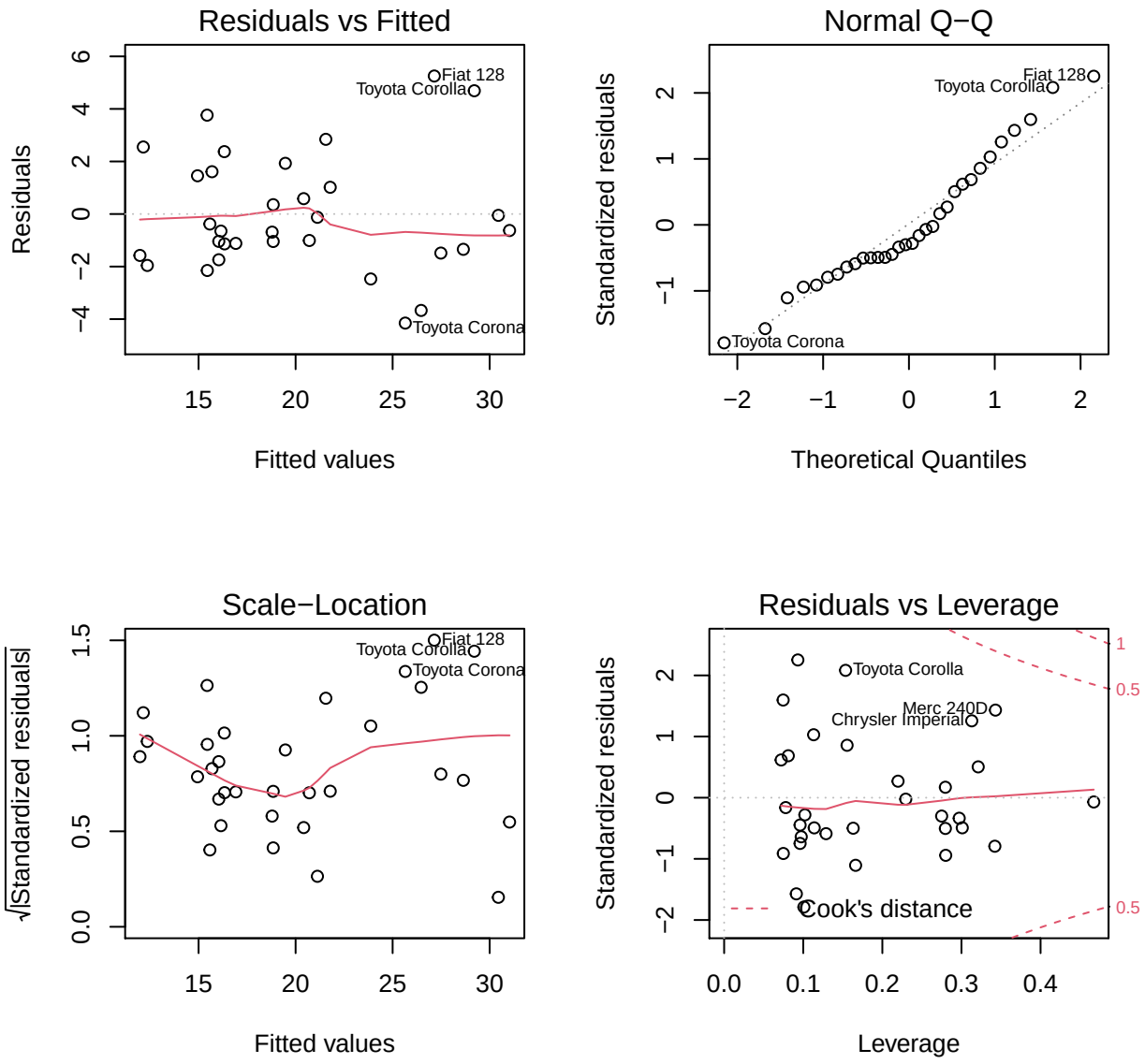


**Figura** 9: Gráfico de resíduos do modelo reduzido 3

## 3.2 Dados reduzidos

O modelo `mod_red1` apresentou o menor valor de R2 ajustado e na análise de resíduo apontou a possível falta de um termo quadrático no modelo. Por essas razões, o modelo foi descartado para nova avaliação.

Em relação aos modelos `mod_red2` e `mod_red3` foi realizado uma nova análise removendo os carros identificados como possíveis *outliers*.

✓ **Modelo reduzido 2**

Os resultados são apresentados na Tabela 8 e na Figura 10 indicam que o formato do motor e o efeito da interação permanecem sendo significativos para explicar a variabilidade presente na eficiência do carro. Além disso, os pontos estão mais próximos da linha tracejada no gráfico **Normal Q-Q**. Entretanto, pelo gráfico **Scale-Location**, há evidências de heterogeneidade de variâncias e outros carros foram identificados como possíveis *outliers*.

```r
mtcars_red <- mtcars %>%
  slice(-8L, -20L, -18L)
mod_red21 <- lm(mpg ~ vs + wt*am, data=mtcars_red)
summary(mod_red21)
```

**Tabela** 8: Estimativas dos parâmetros do modelo reduzido 2 sem os possíveis outliers

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 26.94 | 2.545 | 10.59 | 0.000 |
| vs | 1.87 | 0.859 | 2.18 | 0.040 |
| wt | -2.85 | 0.621 | -4.58 | 0.000 |
| am | 12.34 | 3.074 | 4.01 | 0.001 |
| wt:am | -4.13 | 1.041 | -3.97 | 0.001 |

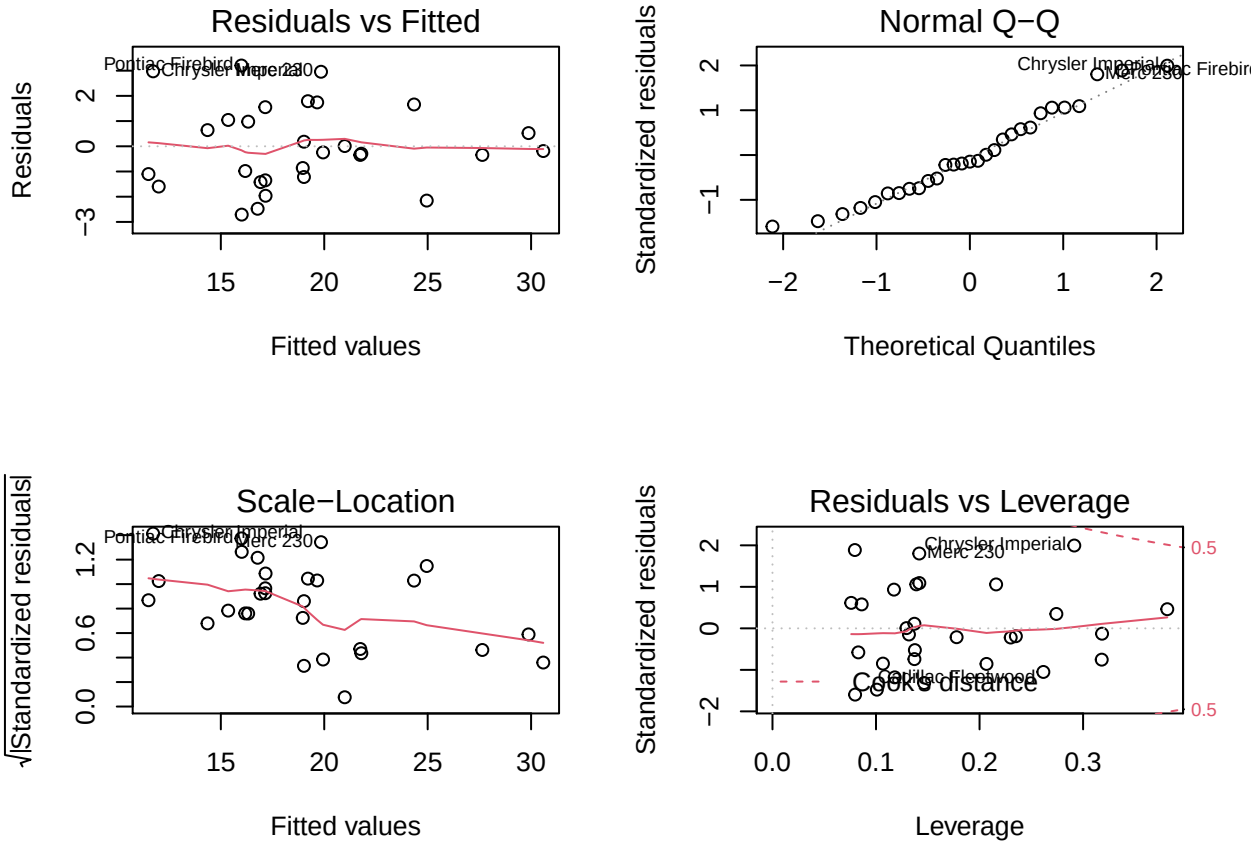


**Figura** 10: Gráfico de resíduos do modelo reduzido 2 sem outliers

✓ **Modelo reduzido 3**

Os resultados são apresentados na Tabela 9 e na Figura 11 indicam que não houve grandes mudanças nos gráficos de resíduos. Porém, houve uma mudança na significância da interação entre `wt` e `cyl8` e também, novos foram identificados como possíveis *outliers*.

```r
mtcars_red1 <- mtcars %>%
  slice(-18L, -20L, -21L)
mod_red31 <- lm(mpg ~ wt*factor(cyl), data=mtcars_red1)
summary(mod_red31)
```

**Tabela** 9: Estimativas dos parâmetros do modelo reduzido 3 sem os possíveis outliers

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 36.06 | 2.61 | 13.816 | 0.000 |
| wt | -4.45 | 1.08 | -4.109 | 0.000 |
| factor(cyl)6 | -7.65 | 7.21 | -1.061 | 0.300 |
| factor(cyl)8 | -12.19 | 3.81 | -3.198 | 0.004 |
| wt:factor(cyl)6 | 1.67 | 2.40 | 0.696 | 0.494 |
| wt:factor(cyl)8 | 2.26 | 1.28 | 1.764 | 0.091 |

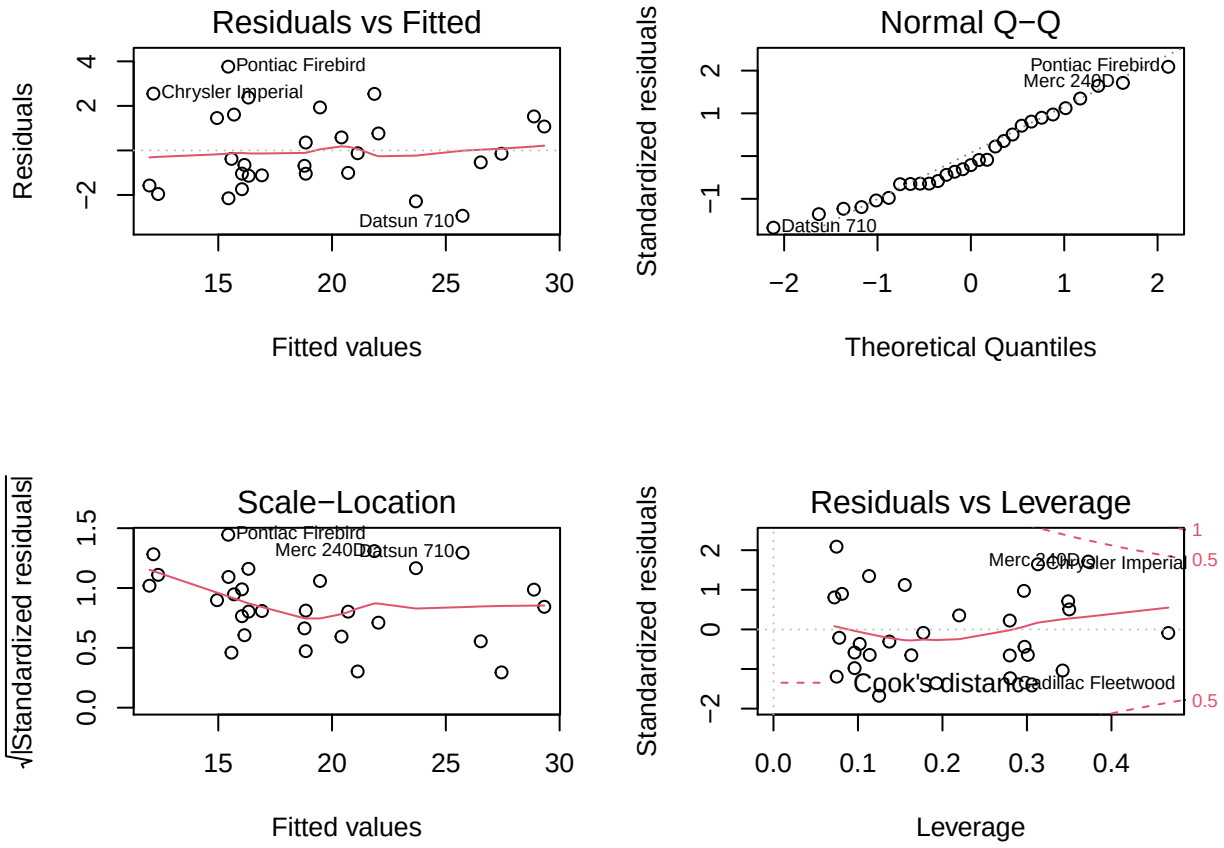


**Figura** 11: Gráfico de resíduos do modelo reduzido 3 sem outliers

# 4 Conclusão e Discussão

Diante do exposto, o modelo mais adequado para explicar a variabiliade presente na eficiência dos carros modelos 1973-1974 é o modelo `mod_red3`. A opção por esse modelo foi em função dos gráficos dos resíduos se manterem com

o mesmo padrão com ou sem os carros `Toyota Corona`, `Toyota Corolla` e `Fiat 128`, apontados como possíveis *outliers*. Esses carros não foram identificados como pontos de alavanca, entretanto, são pontos influentes, pois a significância de um dos parâmetros do modelo foi alterada. Dessa forma, seria interessante levantar mais informações sobre esses carros, antes de removê-los por definitivo do conjunto de dados.

Portanto, o modelo `mod_red3` estimado é dado por

$$\hat{mpg}_i = 39,57 - 5,65wt_i - 11,16cyl6_i - 15,70cyl8_i + 2,87wt_i \times cyl6_i + 3,46wt_i \times cyl8_i, \quad i = 1, \ldots, 32,$$

A partir do modelo estimado, as seguintes interpretações podem ser feitas:

- `wt`: com $p$-valor=0,00o; temos evidências de que a cada 1000 libras que se aumenta no carro há uma redução de -5,65 mpg na eficiência média.
- `cyl6`: com $p$-valor = 0,244; temos indicativos de que não existe diferença significativa entre quatro cilindros (casela de referência) e seis cilindros em relação eficiência média.
- `cyl8`: com $p$-valor = 0,003; temos sinais de que existe diferença significativa entre quatro cilindros (casela de referência) e oito cilindros em relação eficiência média.

Portanto, marginalmente, há o efeito do peso e do número de cilindros

- `wt:cyl6`: com $p$-valor = 0,366; temos evidências de que não existe diferença significativa entre a inclinação de quatro cilindros (casela de referência) e de seis cilindros em relação a eficiência média.
- `wt:cyl8`: com $p$-valor = 0,043; temos evidências de que existe diferença significativa entre a inclinação de quatro cilindros (casela de referência) e de oito cilindros em relação a eficiência média.

Portanto, temos efeito da interação entre peso e número de cilindro do carro. Dessa forma, a hipótese de que modificações na estrutura do automóvel aumenta o seu consumo médio foi validada. Nesse caso, as modificações no peso e no número de cilindros do carro podem explicar a variabilidade presente no consumo médio de gasolina.

Por fim, informações como relação peso e torque poderiam ser utilizadas no lugar de peso e potência. Outra informações que poderia ser utilizada é o tipo de carro, uma vez que carros esportivos são bem diferentes de sedãs.

## Agradecimentos

Ao professor Athos Damiani pelas aulas e dedicação ao curso. Aos mecânicos Marcelo Prataviera e Taka Kurihara e, também, aos alunos do curso de Engenharia Mecânica da UTFPR/Londrina, João Pedro Alves Cordeiro dos Santos e Pedro Henrique Barion pelo auxílio na compreensão das estruturas de um carro.

## Referências bibliográficas

ALMEIDA FILHO, G.M. **Programa INOVAR-AUTO: atendimento das metas de eficiência energética e suas externalidades**. 2018. Dissertação (Mestrado em Ciências) - Universidade de São Paulo, São Paulo.

CRISE petrolífera de 1973. **Wikipedia**. Disponível em: https://pt.wikipedia.org/wiki/Crise_petrol%C3%ADfera_de_1973. Acesso em: 28 de jan. de 2021.

FUEL efficiency. **Wikipedia**. Disponível em: https://en.wikipedia.org/wiki/Fuel_efficiency. Acesso em: 28 de jan. de 2021.

HENDERSON, H.V.; VELLEMAN. P.F. Building multiple regression models interactively. **Biometrics**, v.37, p.391-411, 1981.

HOCKING, R.R. The analysis and selection of variables in linear regression. **Biometrics**, v.32, p.1-49, 1976.

JAMES, G.; WITTEN, D.; HASTIE. T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R**. New York: Springer, 2013.

R Core Team (2020). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

TOP reasons for supporting cleaner, more efficient vehicles. **Global fuel economy initiative**. Disponível em: https://www.globalfueleconomy.org/media/45140/top-reasons-leaflet.pdf. Acesso em: 28 de jan. de 2021.