

# Prova 2

Paulo Ricardo Seganfredo Campana

6 de outubro de 2023

## Questão:

Os dados da **Tabela B.3** do livro do *Montgomery* podem ser encontrados no pacote *MPV* do **R** como o nome *table.b3*. Estes dados consistem de 32 observações sobre o consumo de combustível de diferentes automóveis, com variáveis sobre o desempenho da quilometragem da gasolina e características físicas e/ou de performance.

- a) Ajuste um modelo de regressão múltipla que explica o consumo de combustível do carro  $y$  (em milhas por galão) através da cilindrada ( $x_1$ ), dos cavalos de potência ( $x_2$ ), do comprimento ( $x_8$ ), da largura ( $x_9$ ) e seu peso ( $x_{10}$ ).

```
data <- MPV::table.b3
fit1 <- lm(y ~ x1 + x2 + x8 + x9 + x10, data)
fit1$coefficients
## (Intercept)          x1          x2          x8          x9          x10
##    29.2688    -0.0346    0.0200    0.1433   -0.1869   -0.0045
```

$$\hat{y} = 29.3 - 0.0346x_1 + 0.02x_2 + 0.143x_8 - 0.187x_9 - 0.004x_{10}$$

- b) Através do teste  $F$ , você acha que o modelo adotado é razoável? Justifique sua resposta.

```
summary(fit1)$fstatistic
## F-statistic: 21.8 on 5 and 26 DF,  p-value: 1.54e-08
```

Sim, o teste  $F$  trás p-valor baixo de  $1.54 \times 10^{-8}$  o que indica que a regressão é significativa.

c) O  $R^2$  e  $R_a^2$  sugerem que o modelo proposto explica razoavelmente os dados?

```
summary(fit1)$r.squared
## [1] 0.807
summary(fit1)$adj.r.squared
## [1] 0.77
```

Sim, as variáveis do modelo explicam em torno de 80% da variabilidade do consumo de combustível.

d) Dadas as análises das letras **b)** e **c)** como podemos justificar os resultados dos testes  $t$  para os coeficientes da regressão?

```
summary(fit1) |> broom::tidy() |> knitr::kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	29.269	15.403	1.900	0.069
x1	-0.035	0.022	-1.593	0.123
x2	0.020	0.038	0.527	0.603
x8	0.143	0.091	1.581	0.126
x9	-0.187	0.247	-0.757	0.456
x10	-0.004	0.004	-1.265	0.217

Por mais que pelo teste  $F$  e coeficientes  $R^2$  a regressão seja razoável, nenhum coeficiente do modelo é significativo no teste  $t$ , isso indica que o modelo sofre de multicolinearidade já que a correlação entre as variáveis regressoras são muito altas, o efeito da variável  $x_{10}$  por exemplo, já foi explicado pelas anteriores.

```
cor(fit1$model) |> knitr::kable()
```

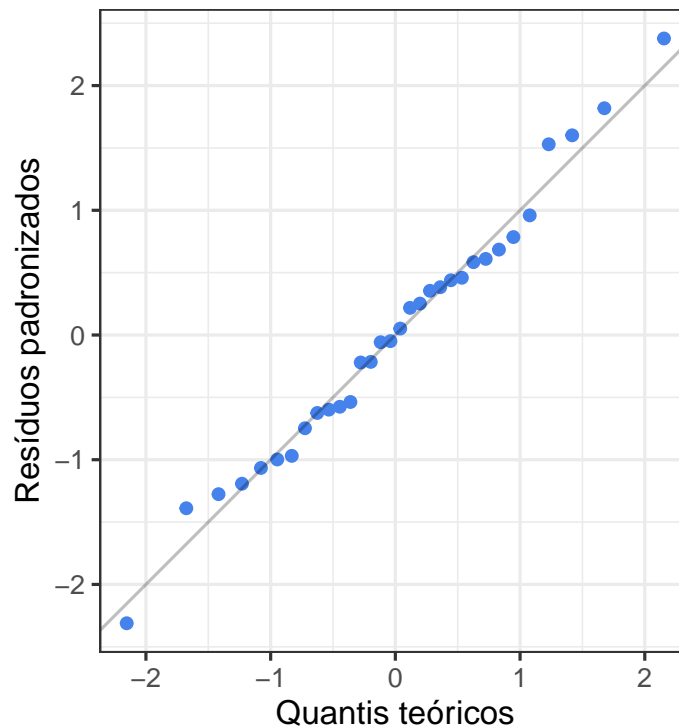
	y	x1	x2	x8	x9	x10
y	1.000	-0.879	-0.807	-0.755	-0.773	-0.863
x1	-0.879	1.000	0.945	0.855	0.801	0.946
x2	-0.807	0.945	1.000	0.797	0.718	0.883
x8	-0.755	0.855	0.797	1.000	0.885	0.948
x9	-0.773	0.801	0.718	0.885	1.000	0.902
x10	-0.863	0.946	0.883	0.948	0.902	1.000

e) Estime agora um modelo de regressão que relaciona o consumo de combustível do carro  $y$  apenas com a cilindrada  $x_1$ . Considerando este modelo faça as seguintes análises.

```
fit2 <- lm(y ~ x1, data)
fit2$coefficients
## (Intercept)      x1
##      33.7227    -0.0474
```

$$\hat{y} = 33.722 - 0.0474x_1$$

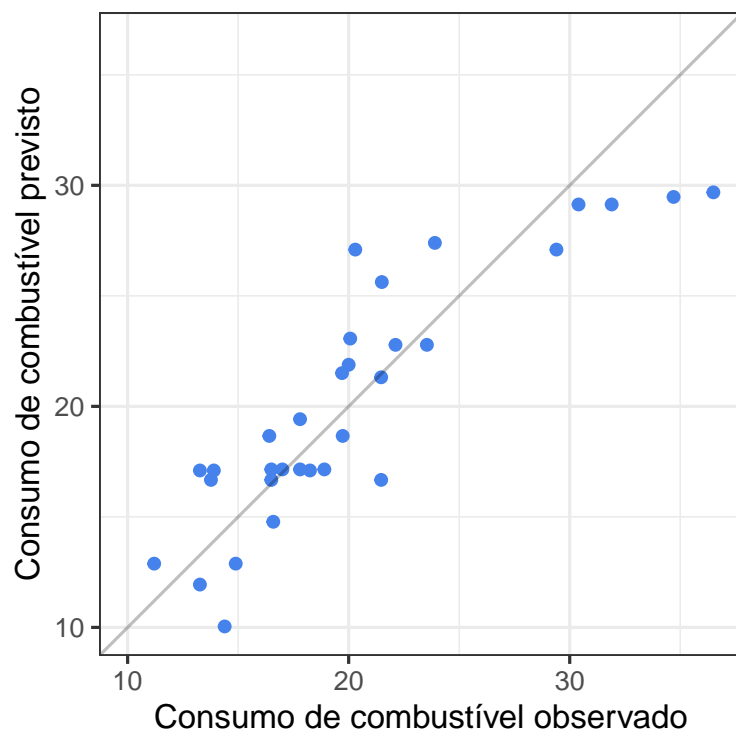
- f) Construa um gráfico para verificar a suposição de normalidade? Parece haver algum problema com esta suposição?



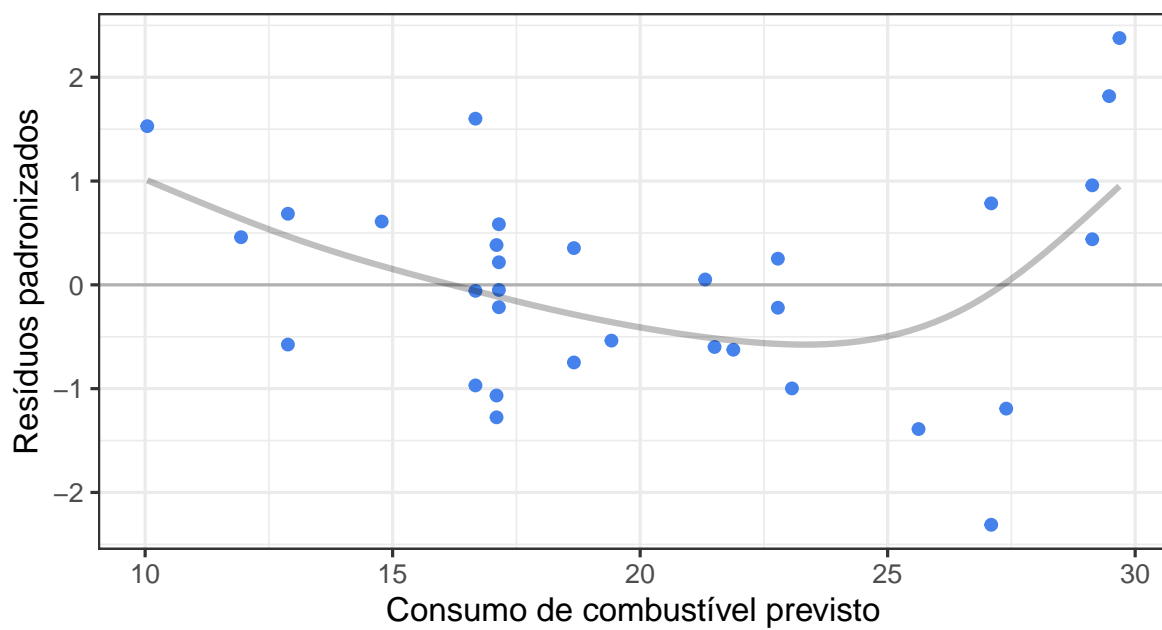
Não, o Q-Q plot não apresenta altos desvios da normalidade.

- g) Construa um gráfico dos valores observados  $y$  versus a resposta prevista. Este gráfico é frequentemente utilizado para verificar qual suposição? Analisando este gráfico, o que é possível observar?

É usado para verificar a linearidade do modelo, neste gráfico vemos que para altos valores do consumo de combustível, o modelo está prevendo abaixo do esperado.



h) Construa um gráfico dos resíduos padronizados versus a resposta prevista. O que é possível analisar neste gráfico?



Vemos que os resíduos não possuem média 0 e variância constante para diferentes valores de  $y$ , isso indica a heterocedasticidade do modelo.

i) Verifique a hipótese de normalidade. Comente os resultados.

```
nortest::lillie.test(rstandard(fit2))  
## D = 0.08, p-value = 0.898  
shapiro.test(rstandard(fit2))  
## W = 1, p-value = 0.955  
moments::jarque.test(rstandard(fit2))  
## JB = 0.2, p-value = 0.922
```

Nenhum teste rejeita a normalidade dos resíduos.

j) Verifique a hipótese de linearidade. Comente os resultados.

```
lmtest::reset(fit2)  
## RESET = 7.0941, df1 = 2, df2 = 28, p-value = 0.00322  
lmtest::raintest(fit2)  
## Rain = 0.38034, df1 = 16, df2 = 14, p-value = 0.967
```

Os testes para a linearidade do modelo discordam em seus resultados.

O teste RESET indica que o modelo pode ser melhorado com transformações não lineares das variáveis regressoras. O teste Rainbow nos diz que o mesmo modelo ajustado com um subconjunto dos dados tem performance similar ao modelo original.

k) Verifique a hipótese de autocorrelação. Comente os resultados.

```
lmtest::dwtest(fit2)  
## DW = 1.6773, p-value = 0.17
```

Segundo o teste de Durbin-Watson, não há autocorrelação no modelo.

l) Verifique a hipótese de homocedasticidade. Comente os resultados.

```
lmtest::gqtest(fit2)  
## GQ = 0.72362, df1 = 14, df2 = 14, p-value = 0.723  
lmtest::bptest(fit2, studentize = FALSE)  
## BP = 4.9086, df = 1, p-value = 0.0267  
lmtest::bptest(fit2, studentize = TRUE)  
## BP = 5.3609, df = 1, p-value = 0.0206
```

Os testes para a heterocedasticidade do modelo discordam em seus resultados.

O teste de Goldfeld-Quandt não detecta heterocedasticidade enquanto que os testes de Breush-Pagan e Koenker conseguem detectar.