MAE0217 - Estatística Descritiva - Lista 5

Natalia Hitomi Koza[1]
Rafael Gonçalves Pereira da Silva[2]
Ricardo Geraldes Tolesano[3]
Rubens Kushimizo Rodrigues Xavier[4]
Rubens Gomes Neto[5]
Rubens Santos Andrade Filho[6]
Thamires dos Santos Matos[7]

Julho de 2021

Sumário

[1]Número USP: 10698432
[2]Número USP: 9009600
[3]Número USP: 10734557
[4]Número USP: 8626718
[5]Número USP: 9318484
[6]Número USP: 10370336
[7]Número USP: 9402940

Capítulo 6

Exercício 5

a)

Podemos definir o seguinte modelo de regressão para os dados:

$$y_{custos} = \alpha + \beta x_{cadeiras}$$

Tendo y como o valor dos custos, e x como o número de cadeiras produzidas; temos que α representa os custos independentes da produção de cadeiras, enquanto β representa o custo de prodção de cada cadeira.

b)

```
dados_ex5 <- read.csv("data/l5-e5.csv")
modelo_e5 <- lm(custos ~ n_cadeiras, data=dados_ex5)

dados_ex5$dy <- dados_ex5$custos - mean(dados_ex5$custos)    # (yi - Y)

dados_ex5$dx <- dados_ex5$n_cadeiras - mean(dados_ex5$n_cadeiras) # (xi - X)

dados_ex5$dx2 <- dados_ex5$dx ** 2                    # (xi - X)^2

dados_ex5$dxy <- dados_ex5$dx * dados_ex5$dy           # (xi - X) * (yi - Y)

beta <- sum(dados_ex5$dxy) / sum(dados_ex5$dx2)
alpha <- mean(dados_ex5$custos) - (beta * mean(dados_ex5$n_cadeiras))
```

Para extrapolar os custos de produzir 200 cadeiras podemos calcular:

```
custos_200 <- alpha + (beta * 200)
custos_200
```

```
## [1] 2095.778
```

c)

Para encontrar o número de cadeiras que precisam ser vendidas temos que encontrar:

$$\begin{aligned} 20n &\geq \alpha + \beta n \\ 20n - \beta n &\geq \alpha \\ (20 - \beta)n &\geq \alpha \\ n &\geq \frac{\alpha}{20 - \beta} \end{aligned}$$

Assim podemos encontrar que:

```
n_lucro <- alpha / (20 - beta)
n_lucro
```

```
## [1] 81.4829
```

Sendo assim temos um número de cadeiras de 82.

Exercício 8

Formatamos a tabela de forma a facilitar a análise. As variáveis foram renomeadas de acordo com o seguinte dicionário:

- consumo_oxigenio_pico - consumo de oxigênio no pico do exercício em ml/kg/min
- sexo - F: feminino, M: masculino
- idade - idade do paciente em anos
- peso - peso do paciente em kg
- classificacao_nyha - classe funcional pelo critério NYHA (1 a 4)
- carga_esteira - carga utilizada na esteira ergométrica
- frequencia_cardiaca - frequência cardíaca em batimentos por minuto
- razao_troca_respiratoria - razão de troca respiratória em VCO2/VO2

Mostrando algumas linhas da tabela:

```
esforco <- read_excel("data/esforco_6_18.xlsx", na=".")
pander(head(esforco, 10), caption="Dados contidos no arquivo esforco")
```

Tabela 1: Dados contidos no arquivo esforco (continued below)

sexo	idade	peso	classificacao_nyha	carga_esteira	frequencia_cardiaca
M	38	54	2	71	118
M	49	80	1	91	113
F	65	56	2	37	148
M	52	78	2	127	144
F	52	59	4	43	107
F	58	62	1	60	135
F	24	42	3	32	117
F	39	55	2	63	147
F	48	77	3	71	175
M	50	81	1	112	148

consumo_oxigenio_pico	razao_troca_respiratoria
14.1	1.26
16.3	1.09
9.9	1.1

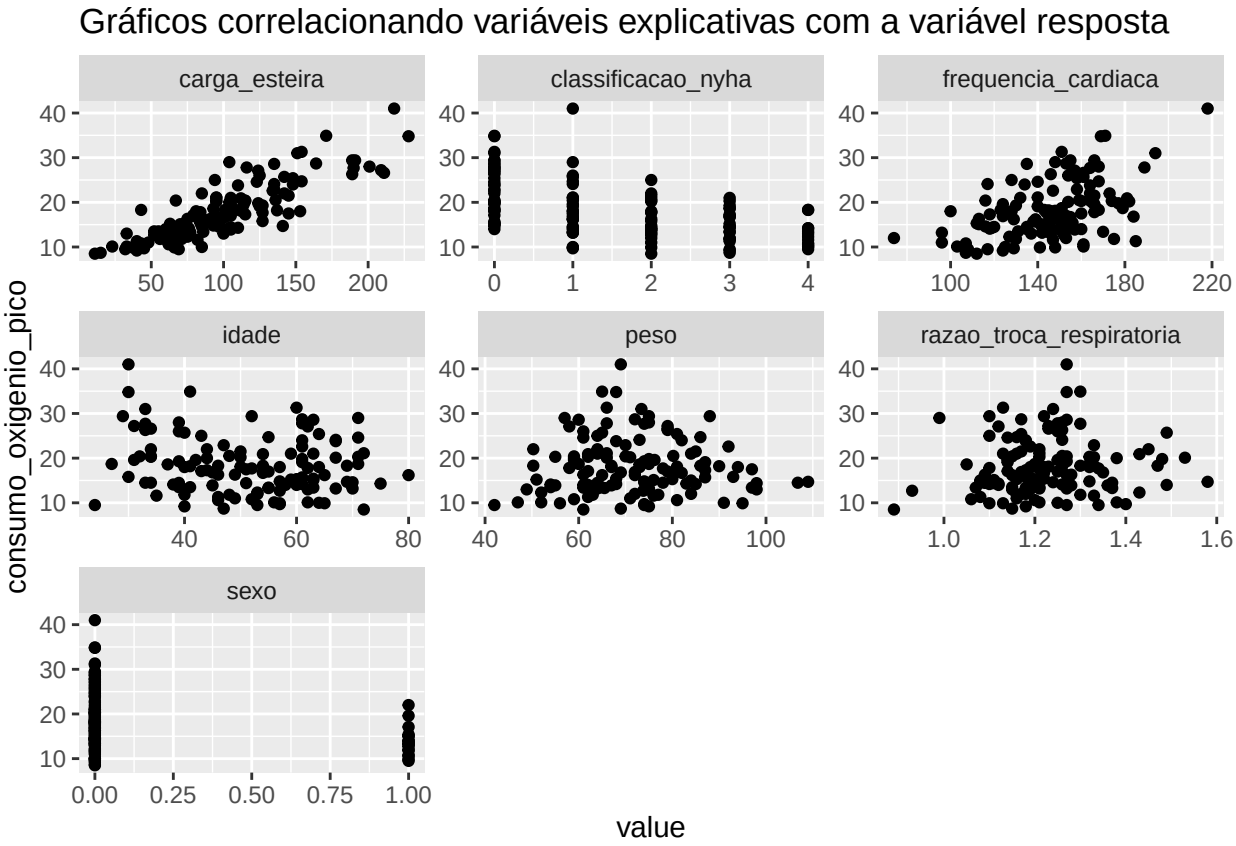
consumo_oxigenio_pico	razao_troca_respiratoria
17.7	1.34
10.8	1.06
14	1.12
9.5	1.27
13.9	1.28
11.8	1.16
18.1	1.23

```
summary(esforco)
```

```
##      sexo               idade            peso
##  Length:127         Min.   :24.00   Min.   : 42.00
##  Class :character   1st Qu.:40.50   1st Qu.: 62.00
##  Mode  :character   Median :53.00   Median : 72.00
##                     Mean   :51.54   Mean   : 71.76
##                     3rd Qu.:61.00   3rd Qu.: 80.00
##                     Max.   :80.00   Max.   :109.00
##
##  classificacao_nyha carga_esteira   frequencia_cardiaca
##  Min.   :0.000      Min.   : 11.0   Min.   : 74.0
##  1st Qu.:0.000      1st Qu.: 70.0   1st Qu.:129.5
##  Median :1.000      Median : 97.0   Median :147.0
##  Mean   :1.551      Mean   :101.5   Mean   :144.9
##  3rd Qu.:3.000      3rd Qu.:125.0   3rd Qu.:161.0
##  Max.   :4.000      Max.   :228.0   Max.   :218.0
##                     NA's   :2
##  consumo_oxigenio_pico razao_troca_respiratoria
##  Min.   : 5.20         Min.   :0.890
##  1st Qu.:13.50         1st Qu.:1.150
##  Median :17.10         Median :1.210
##  Mean   :18.06         Mean   :1.220
##  3rd Qu.:21.05         3rd Qu.:1.275
##  Max.   :41.00         Max.   :1.580
##
```

Converteremos a variável sexo para os valores 0 se masculino e 1 se feminino. Construindo os gráficos de dispersão correlacionando as variáveis explicativas com a variável resposta:

```
esforco <- read_excel("data/esforco_6_18.xlsx", na=".")
esforco <- drop_na(esforco)
esforco$sexo <- as.integer(esforco$sexo == "F")
esforco %>% gather(-consumo_oxigenio_pico, key="key", value="value") %>% ggplot(aes(x=value, y=consumo_
```



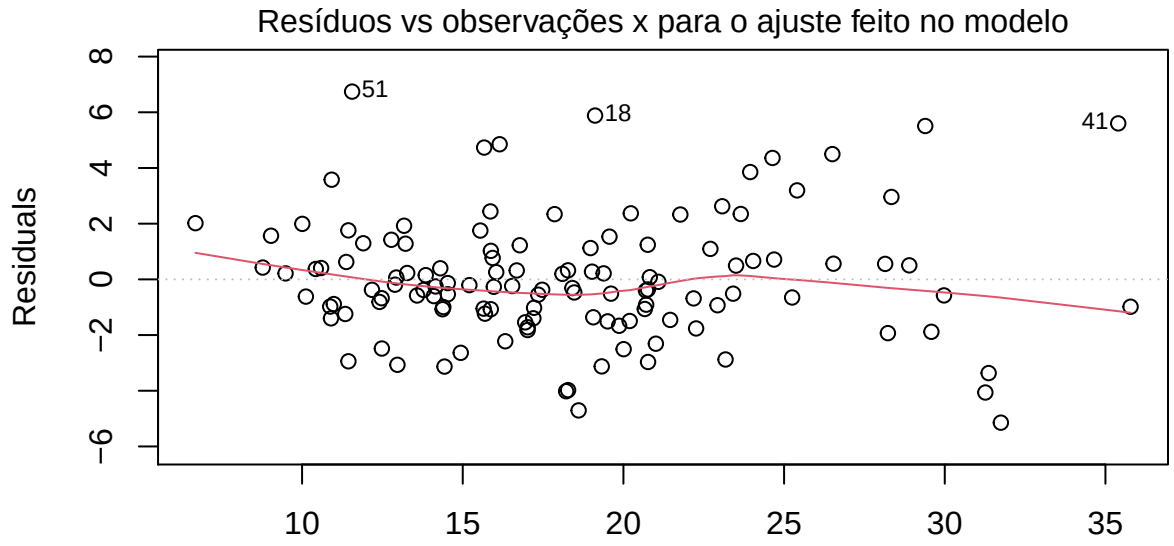
Adotaremos o modelo de regressão linear: y_i = a + b * sexo + c * idade + d * peso + e * classificacao_nyha + i * carga_esteira + f * frequencia_cardiaca + g * razao_troca_respiratoria + e_i Onde a é o intercepto, $[c, d, e, i, f, g]$ são coeficientes, e_i são erros aleatórios não correlacionados, e as outras variáveis são explicadas pelo dicionário acima.

Ajustando o modelo e apresentando erros padrões, gráficos e outras informações:

```
fit_titles <- list("Resíduos vs observações x para o ajuste feito no modelo",
                   "Gráfico Q-Q normal para o ajuste feito no modelo",
                   "Resíduos normalizados vs observações x para o ajuste feito no modelo",
                   "Resíduos normalizados vs influência das observações para o ajuste feito no modelo")

mostrarAjuste <- function(dados, ajuste) {

  print(summary(ajuste))
  plot(ajuste,
       caption=fit_titles)

  #confidence_intervals <- confint(ajuste)
  #k <- kable(confidence_intervals, caption="Intervalos de confiança para o ajuste dos parâmetros do mo
  #print(k)

  return(ajuste)
}
```

```r
ajuste <- lm(consumo_oxigenio_pico ~ sexo + idade + peso + classificacao_nyha + carga_esteira + frequen
mostrarAjuste(esforco, ajuste)
```

```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ sexo + idade + peso + classificacao_nyha +
##     carga_esteira + frequencia_cardiaca + razao_troca_respiratoria,
##     data = esforco)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.146 -1.232 -0.267  1.094  6.742
##
## Coefficients:
##                            Estimate Std. Error t value
## (Intercept)              25.772135   3.368114   7.652
## sexo                     -1.849114   0.650137  -2.844
## idade                     0.020785   0.019329   1.075
## peso                     -0.192197   0.018021 -10.665
## classificacao_nyha       -0.678217   0.211915  -3.200
## carga_esteira             0.122720   0.007999  15.343
## frequencia_cardiaca       0.034966   0.010168   3.439
## razao_troca_respiratoria -9.002521   1.851729  -4.862
##                           Pr(>|t|)
## (Intercept)              6.16e-12 ***
## sexo                      0.00526 **
## idade                     0.28442
## peso                      < 2e-16 ***
## classificacao_nyha        0.00177 **
## carga_esteira             < 2e-16 ***
## frequencia_cardiaca       0.00081 ***
## razao_troca_respiratoria 3.66e-06 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.254 on 117 degrees of freedom
## Multiple R-squared:  0.8802, Adjusted R-squared:  0.873
## F-statistic: 122.8 on 7 and 117 DF,  p-value: < 2.2e-16
```


Resíduos vs observações x para o ajuste feito no modelo
Residuals
51
18
41
8
6
4
2
0
−2
−6
10
15
20
25
30
35
Fitted values
lm(consumo_oxigenio_pico ~ sexo + idade + peso + classificacao_nyha + carga ...

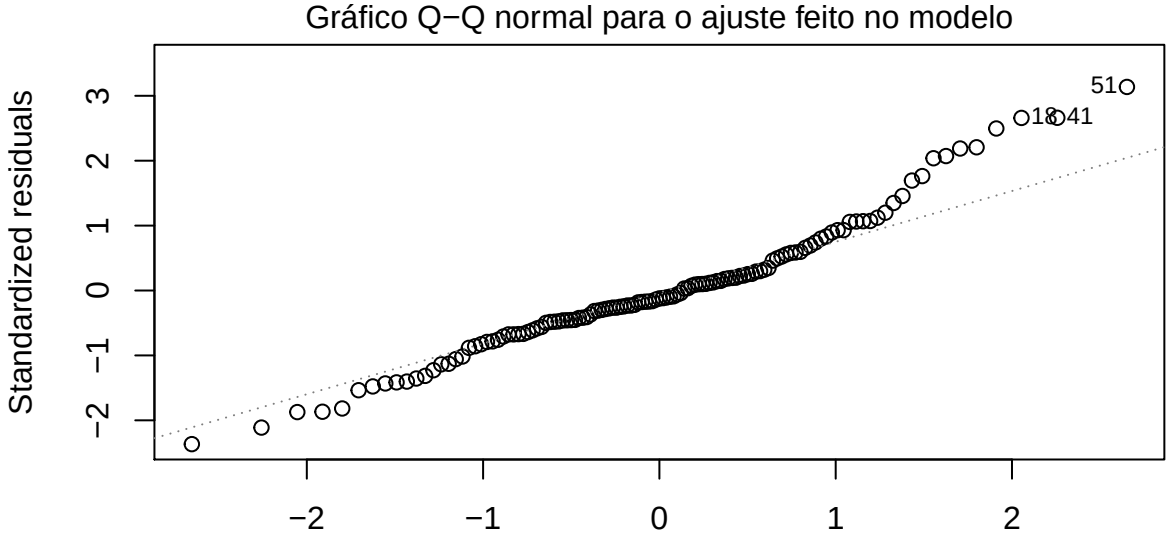
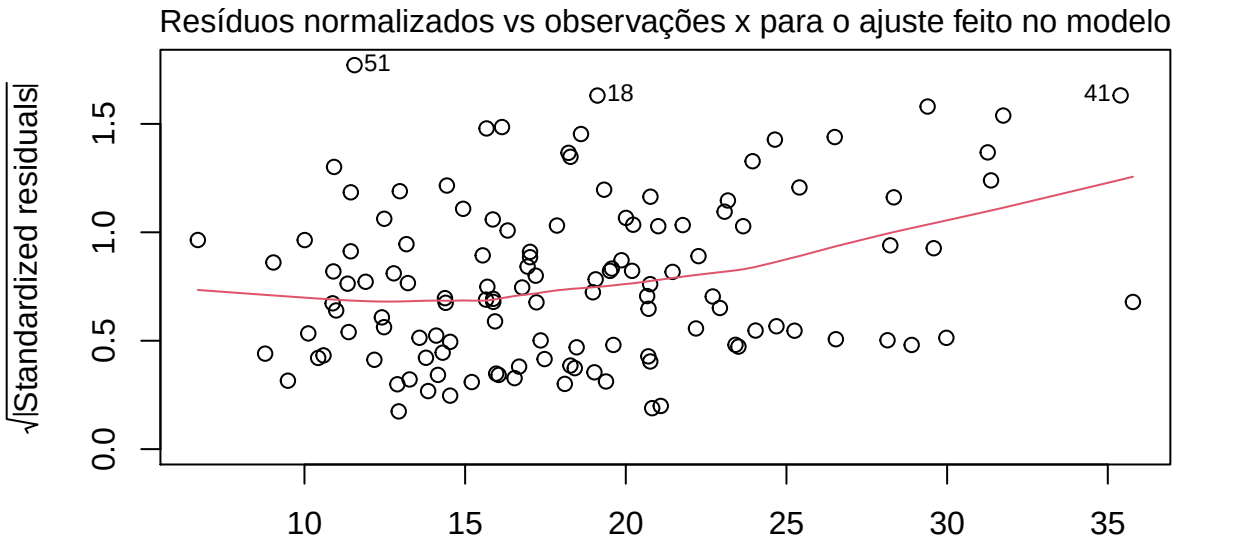
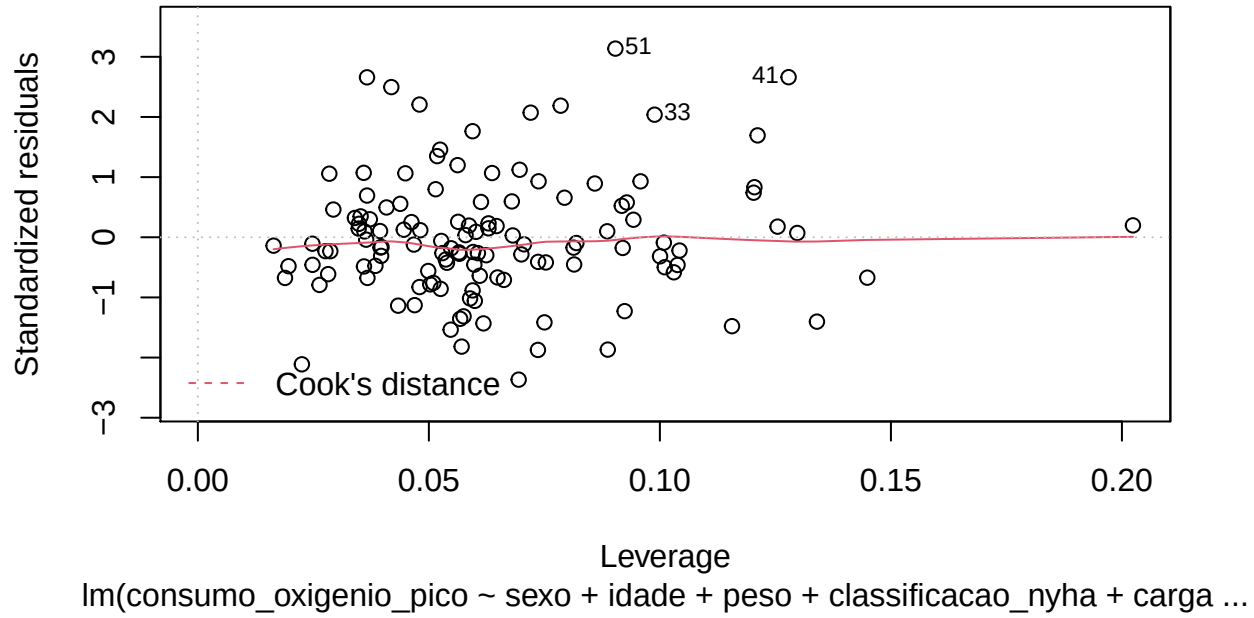


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
51
18
41
3
2
1
0
−1
−2
−2
−1
0
1
2
Theoretical Quantiles
lm(consumo_oxigenio_pico ~ sexo + idade + peso + classificacao_nyha + carga ...






```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ sexo + idade + peso + classificacao_nyha +
##     carga_esteira + frequencia_cardiaca + razao_troca_respiratoria,
##     data = esforco)
##
## Coefficients:
##              (Intercept)                      sexo
##                 25.77214                   -1.84911
##                    idade                      peso
##                  0.02079                   -0.19220
##       classificacao_nyha             carga_esteira
##                 -0.67822                    0.12272
```

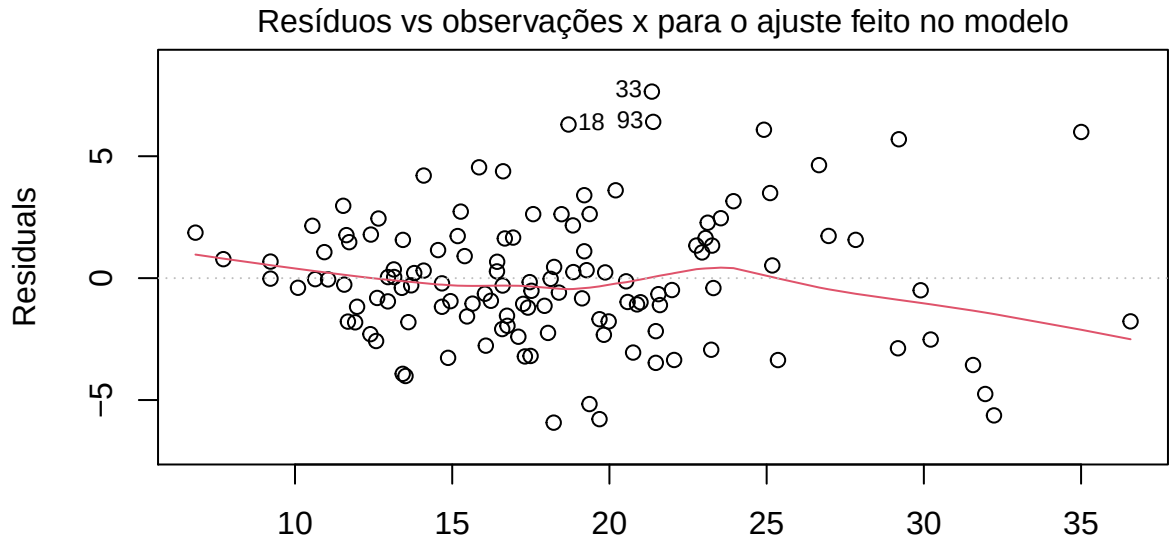
```
##      frequencia_cardiaca  razao_troca_respiratoria
##                  0.03497                  -9.00252
```

O valor de R^2 e o gráfico de resíduos indicam o esperado: que apesar de algumas variáveis se correlacionarem de forma linear visualmente, elas estão organizadas em uma faixa larga, ou seja, a correlação não é precisa. O gráfico QQ normal desvia não se aproxima muito da reta diagonal na ponta direita, o que indica que podemos obter um ajuste melhor.

Queremos utilizar somente os parâmetros mais relevantes no modelo. Utilizaremos somente os parâmetros intercept, sexo, peso e carga_esteira, que possuem os menores valores de Pr(>|t|). Ajustando o modelo:

```
ajuste2 <- lm(consumo_oxigenio_pico ~ sexo + peso + carga_esteira, data=esforco)
mostrarAjuste(esforco, ajuste2)
```

```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ sexo + peso + carga_esteira,
##     data = esforco)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9271 -1.7734 -0.2146  1.6269  7.6504
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.199808   1.420422  12.109  < 2e-16 ***
## sexo          -0.549299   0.711049  -0.773    0.441
## peso          -0.180401   0.020046  -8.999 3.88e-15 ***
## carga_esteira  0.138776   0.006133  22.629  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.666 on 121 degrees of freedom
## Multiple R-squared:  0.8266, Adjusted R-squared:  0.8223
## F-statistic: 192.2 on 3 and 121 DF,  p-value: < 2.2e-16
```


Resíduos vs observações x para o ajuste feito no modelo
Residuals
Fitted values
lm(consumo_oxigenio_pico ~ sexo + peso + carga_esteira)
33
18
93

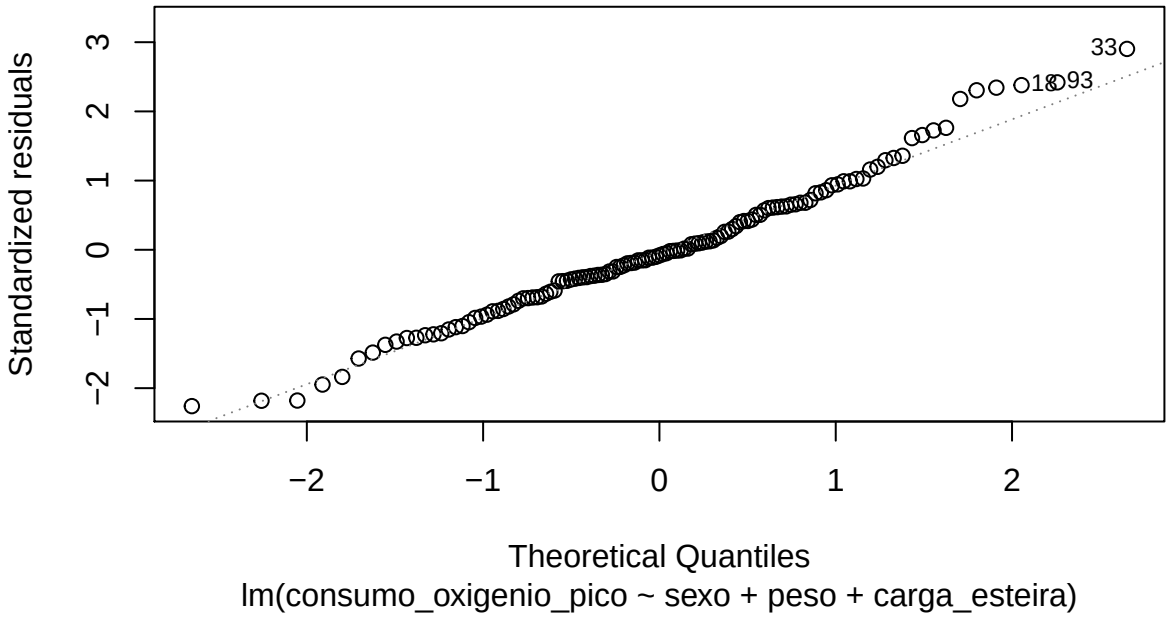
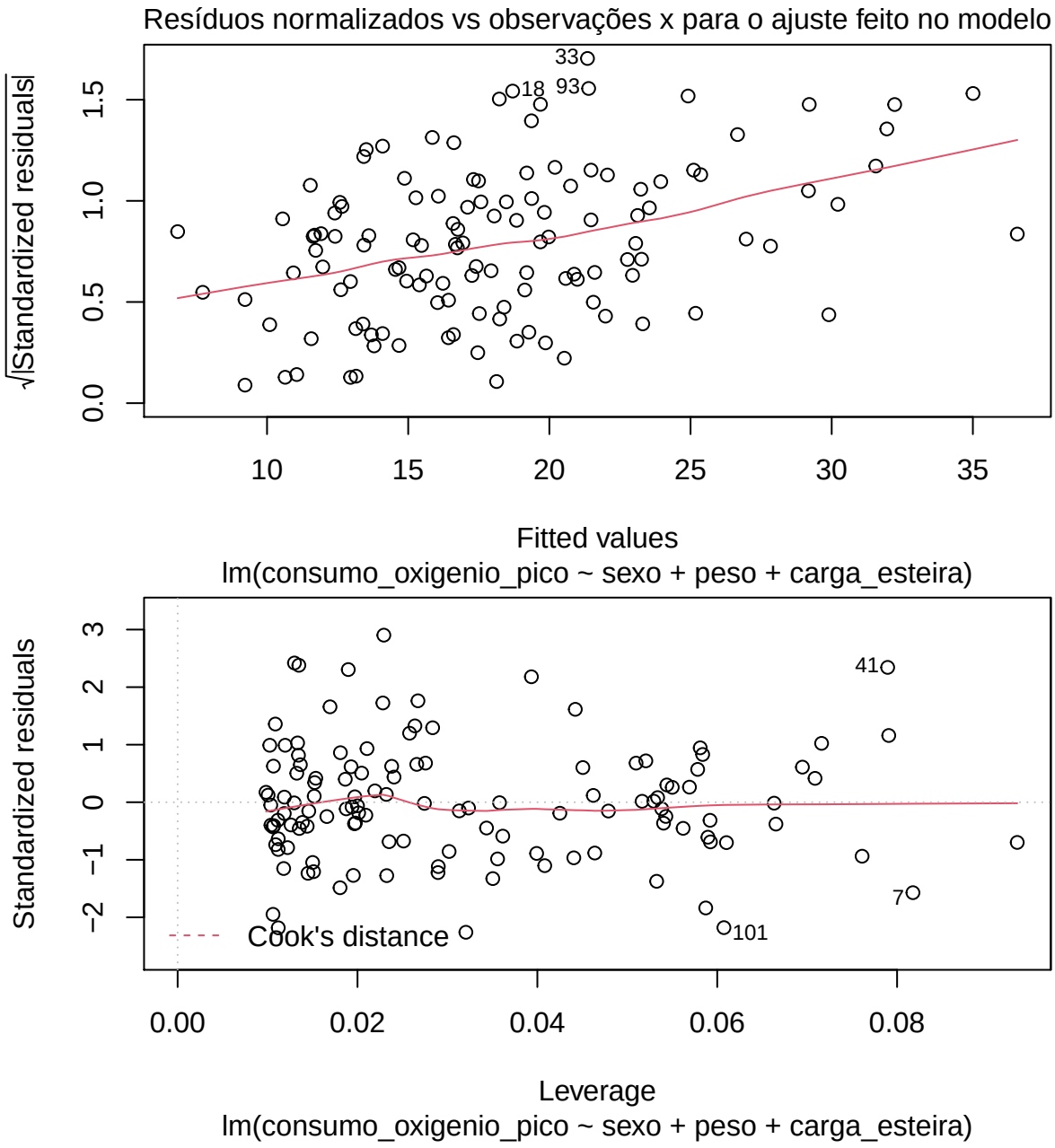


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
Theoretical Quantiles
lm(consumo_oxigenio_pico ~ sexo + peso + carga_esteira)
33
18
93



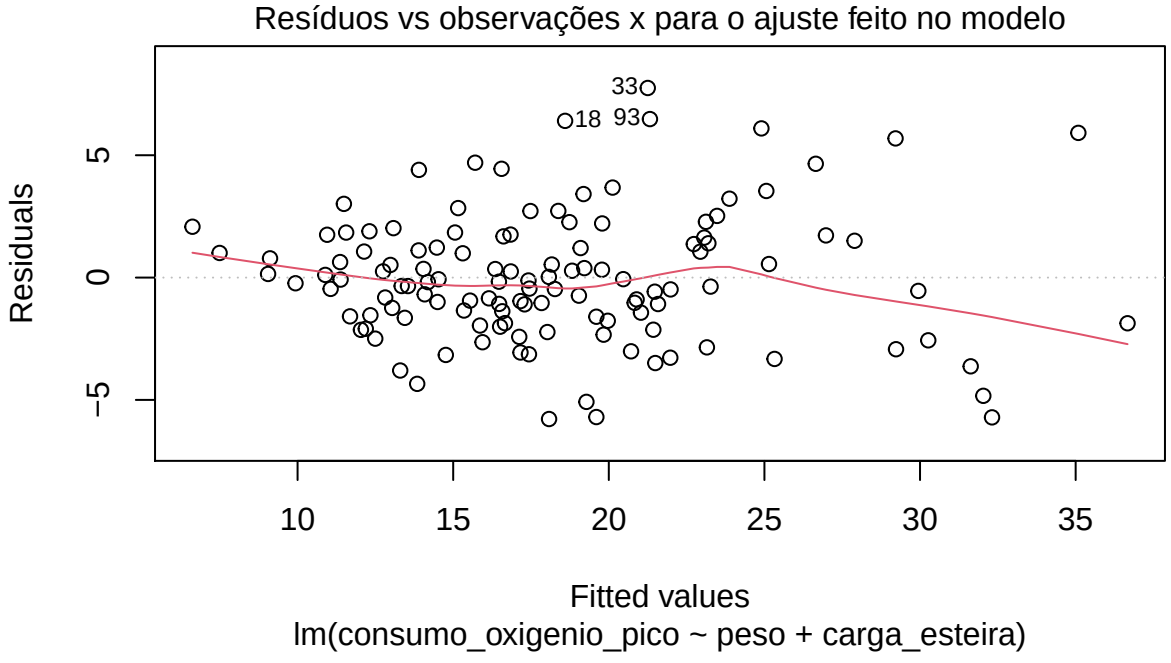

```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ sexo + peso + carga_esteira,
##     data = esforco)
##
## Coefficients:
##   (Intercept)           sexo           peso  carga_esteira
##       17.1998        -0.5493        -0.1804         0.1388
```

Após o novo ajuste com somente as variáveis mais significativas, pode-se observar que o resíduo na verdade aumentou ligeiramente. Entretanto, o valor R^2 diminuiu e os pontos no gráfico Q-Q Normal estão mais próximos da linha diagonal, indicando melhor qualidade do ajuste.

O parâmetro sexo não é mais tão significativo no modelo. Ajustando o modelo sem esse parâmetro:

```
ajuste3 <- lm(consumo_oxigenio_pico ~ peso + carga_esteira, data=esforco)
mostrarAjuste(esforco, ajuste3)
```

```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ peso + carga_esteira, data = esforco)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7810 -1.6461 -0.1862  1.6333  7.7485
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   16.881164   1.356967  12.440  < 2e-16 ***
## peso          -0.179161   0.019948  -8.981 4.05e-15 ***
## carga_esteira  0.140217   0.005832  24.041  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.662 on 122 degrees of freedom
## Multiple R-squared:  0.8257, Adjusted R-squared:  0.8229
## F-statistic:   289 on 2 and 122 DF,  p-value: < 2.2e-16
```



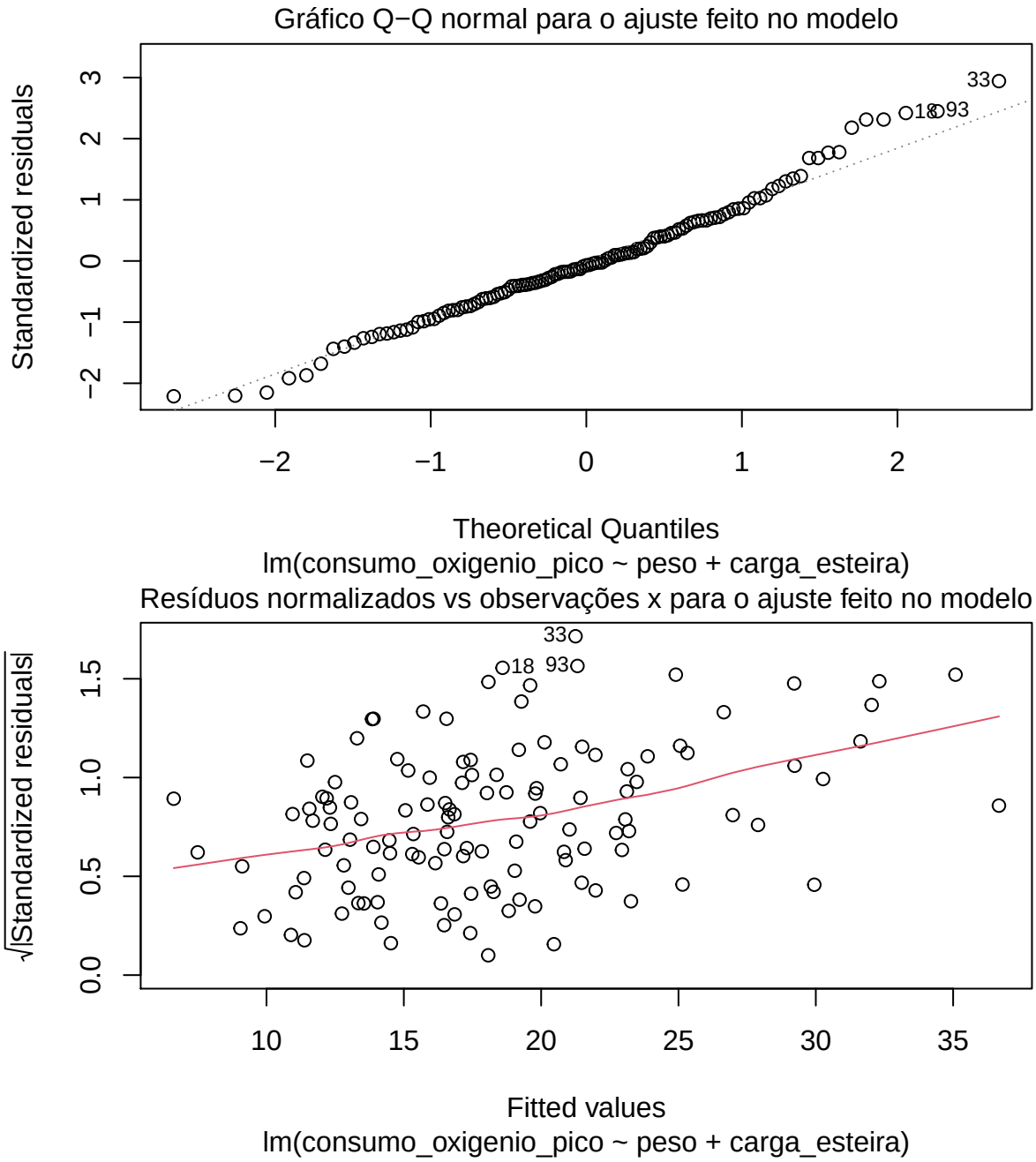
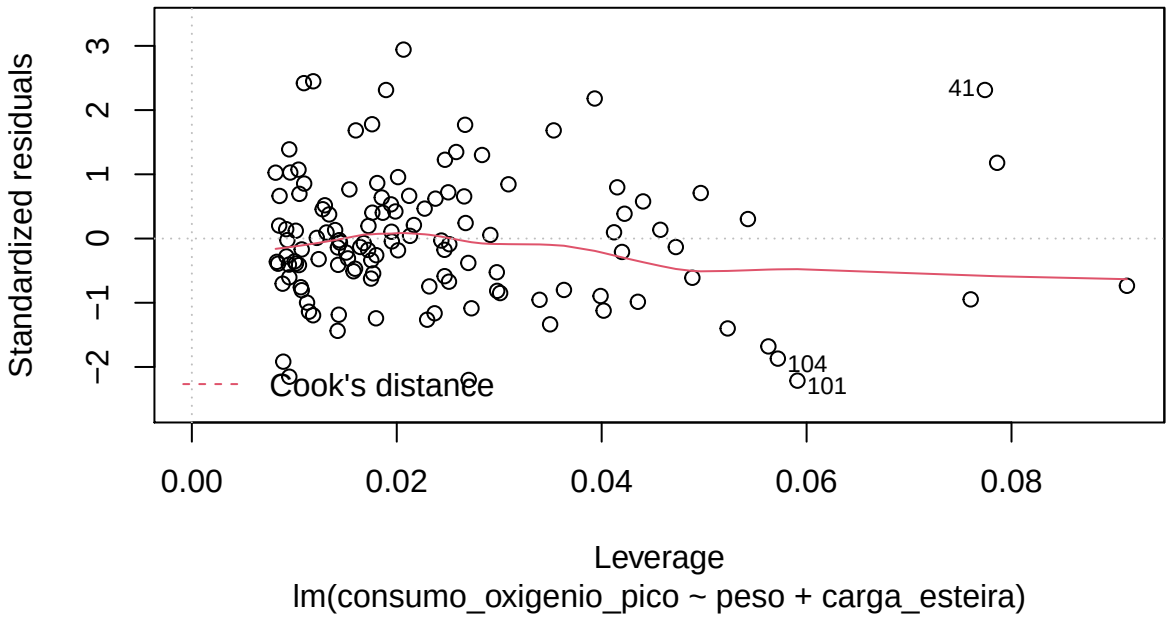


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
33
18
93
−2
−1
0
1
2
3
Theoretical Quantiles
lm(consumo_oxigenio_pico ~ peso + carga_esteira)

Resíduos normalizados vs observações x para o ajuste feito no modelo
√|Standardized residuals|
33
18
93
0.0
0.5
1.0
1.5
10
15
20
25
30
35
Fitted values
lm(consumo_oxigenio_pico ~ peso + carga_esteira)




```
##
## Call:
## lm(formula = consumo_oxigenio_pico ~ peso + carga_esteira, data = esforco)
##
## Coefficients:
##   (Intercept)           peso  carga_esteira
##       16.8812        -0.1792         0.1402
```

A remoção da variável 'sexo' não causou mudança significativa em relação ao ajuste anterior.

Exercício 18

Exercício 19

Partindo da função (6.29), e dado que $P(Y_i = 0|X = x) = 1 - P(Y_i = 1|X = x)$, podemos demonstrar que:

$$
\begin{aligned}
\log \frac{P(Y_i=1|X=x)}{P(Y_i=0|X=x)} &= \alpha + \beta x_i \\
\exp\left(\log \frac{P(Y_i=1|X=x)}{P(Y_i=0|X=x)}\right) &= \exp(\alpha+\beta x_i) \\
\frac{P(Y_i=1|X=x)}{P(Y_i=0|X=x)} &= \exp(\alpha+\beta x_i) \\
\frac{P(Y_i=1|X=x)}{1-P(Y_i=1|X=x)} &= \exp(\alpha+\beta x_i) \\
P(Y_i=1|X=x) &= \exp(\alpha+\beta x_i)(1-P(Y_i=1|X=x)) \\
P(Y_i=1|X=x) &= \exp(\alpha+\beta x_i) - P(Y_i=1|X=x)\exp(\alpha+\beta x_i) \\
P(Y_i=1X=x) + P(Y_i=1|X=x)\exp(\alpha+\beta x_i) &= \exp(\alpha+\beta x_i) \\
P(Y_i=1X=x)(1+\exp(\alpha+\beta x_i)) &= \exp(\alpha+\beta x_i) \\
P(Y_i=1X=x) &= \frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)} \quad \square
\end{aligned}
$$

Assim podemos ver que de fato (6.29) é equivalente a (6.30). Para além disso podemos demonstrar que $0 \leq P(Y_i=1|X=x) \leq 1$, uma vez que:

$$
\begin{aligned}
P(Y_i=1|X=x) = \frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)} &\leq 1 \\
\exp(\alpha+\beta x_i) &\leq 1+\exp(\alpha+\beta x_i) \\
\exp(\alpha+\beta x_i) - \exp(\alpha+\beta x_i) &\leq 1 \\
0 &\leq 1 \quad \square
\end{aligned}
$$

$$
\begin{aligned}
P(Y_i=1|X=x) = \frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)} &\geq 0 \\
\exp(\alpha+\beta x_i) &\geq 0 \\
1 &\geq \frac{0}{\exp(\alpha+\beta x_i)} \\
1 &\geq 0 \quad \square
\end{aligned}
$$

Exercício 21

Capítulo 7

Exercício 1

Exercício 2

Exercício 6