

# Resolução Lista I

Nathalia Gabriella Ferreira dos Santos

2023-09-06

## Pacotes

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.4.2      v purrr   1.0.1
## v tibble  3.2.1      v dplyr  1.1.1
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.4      v forcats 1.0.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

## Questão 1

### Enunciado:

Considere o conjunto de dados abaixo referente a um valor mínimo pago de seguros individuais (em dólares americanos) por danos com acidentes pessoais. Além do valor pago ao segurado, serão consideradas as seguintes variáveis explicativas:

- *legrep*: representação legal (não ou sim)
- *optime*: tempo operacional (em horas) para pagamento do seguro.

```
rm(list = ls())

df <- data.frame(
  valor = c(9.17, 6.79, 16.64, 8.81, 15.86, 10.07, 12.69, 6.91, 5.43, 10.88, 8.37, 11.9, 16.12, 8.74, 9.35),
  legrep = c("nao", "nao", "nao", "nao", "sim", "nao", "sim", "nao", "nao", "sim", "nao", "sim", "sim", "nao", "nao"),
  optime = c(18.2, 13.1, 32.9, 17.4, 29.5, 19.5, 22.7, 13.3, 10.5, 19.6, 16.3, 21.2, 29.7, 17, 18.3)
) %>% as_tibble()

df

## # A tibble: 15 x 3
##   valor legrep optime
##   <dbl> <chr>   <dbl>
## 1  9.17 nao     18.2
## 2  6.79 nao     13.1
## 3 16.6  nao     32.9
## 4  8.81 nao     17.4
## 5 15.9  sim      29.5
```

```
## 6 10.1 nao      19.5
## 7 12.7 sim      22.7
## 8 6.91 nao      13.3
## 9 5.43 nao      10.5
## 10 10.9 sim     19.6
## 11 8.37 nao     16.3
## 12 11.9 sim     21.2
## 13 16.1 sim     29.7
## 14 8.74 nao     17
## 15 9.35 nao     18.3
```

a)

```
x <- df %>% select(legrep, optime) %>%
  mutate(aux = 1) %>%
  pivot_wider(names_from = legrep,
              names_prefix = "legrep_",
              values_from = aux,
              values_fill = 0)
```

x

```
## # A tibble: 15 x 3
##   optime legrep_nao legrep_sim
##   <dbl>   <dbl>   <dbl>
## 1  18.2         1         0
## 2  13.1         1         0
## 3  32.9         1         0
## 4  17.4         1         0
## 5  29.5         0         1
## 6  19.5         1         0
## 7  22.7         0         1
## 8  13.3         1         0
## 9  10.5         1         0
## 10 19.6         0         1
## 11 16.3         1         0
## 12 21.2         0         1
## 13 29.7         0         1
## 14 17          1         0
## 15 18.3         1         0
```

b)

```
fit <- lm(formula = valor ~ legrep + optime, data = df)
```

```
resposta = fit %>% coefficients() %>% cbind()
colnames(resposta) <- "Coeficientes"
row.names(resposta) <- c("b0", "b1", "b2")
resposta
```

```
##   Coeficientes
## b0    0.2205055
## b1    1.0238336
## b2    0.4990082
```

c)

Sim, todas variáveis explicativas são significativas pois  $p - \text{valor} < \alpha = 0,05$  para todo p-valor encontrado.

```
resumo <- summary(fit)
resumo$coefficients[,4]
```

```
## (Intercept)    legrepsim      optime
## 2.272168e-02 7.461696e-10 1.887332e-19
```

d)

```
fit %>% fitted()
```

```
##          1          2          3          4          5          6          7          8
## 9.302455  6.757513 16.637875  8.903248 15.965081  9.951165 12.571825  6.857314
##          9         10         11         12         13         14         15
## 5.460091 11.024900  8.354339 11.823313 16.064882  8.703645  9.352355
```

e)

Sobre a hipótese nula de que os dados seguem distribuição normal contra a hipótese alternativa que os dados não seguem distribuição normal, dado um  $p\text{-valor} > \alpha = 0,05$ , não rejeita-se  $H_0$ , logo, há evidências que os resíduos seguem distribuição normal.

```
fit %>% residuals() %>% shapiro.test()
```

```
##
## Shapiro-Wilk normality test
##
## data:  .
## W = 0.93269, p-value = 0.2992
```

f)

```
X <- x %>% as.matrix()
t(X) %*% X %>% # transposta * matriz normal
  solve()      # inversa da matriz

##          optime  legrep_ao  legrep_sim
## optime      0.00238131 -0.04203011 -0.05843734
## legrep_ao -0.04203011  0.84183151  1.03141900
## legrep_sim -0.05843734  1.03141900  1.63405225
```

g)

```
r2_adj <- resumo$adj.r.squared
```

Dado um  $R^2 = 99.9306984$ , conclui-se que o modelo ficou muito bem ajustado para esse conjunto de dados.