

Exercícios resolvidos - CAP 5 - Modelos de Regressão Linear Simples

Iara Denise Endruweit Battisti, Erikson Kaszubowski, Felipe Micaíl da Silva Smolski

Exercícios

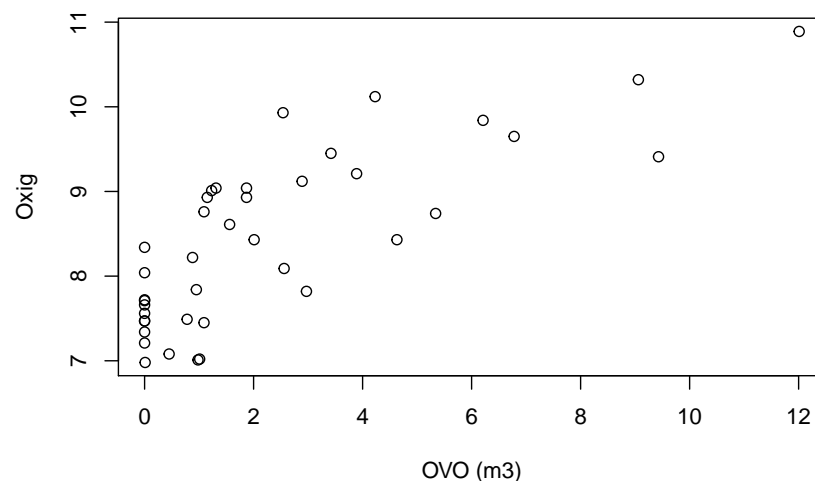
1. No site do livro (<https://smolski.github.io/softwarelivrer/livro.html>) está disponível uma planilha de dados com o nome “peixes1”, na primeira coluna consta a quantidade de ovos (m^3) e na segunda coluna consta a quantidade de oxigênio no rio. O objetivo da pesquisa é comparar alguns ambientes do rio sobre a desova e o crescimento das larvas de peixes. Analisar a relação da quantidade de ovos com a quantidade de oxigênio no rio. Para isso utilize coeficiente de correlação linear e regressão linear simples.

```
library(readxl)
url <- "https://smolski.github.io/softwarelivrer/peixes1.xls"
destfile <- "peixes1.xls"
curl::curl_download(url, destfile)
peixes1 <- read_excel(destfile)

attach(peixes1)
head(peixes1)
```

```
# A tibble: 6 x 2
  `OVO (m3)`  Oxig
    <dbl> <dbl>
1      6.21  9.84
2      4.23 10.1
3      5.34  8.74
4      1.31  9.04
5      3.89  9.21
6      2.01  8.43
```

```
plot(`OVO (m3)`,Oxig)
```



```
cor(`OVO (m3)`,Oxig)
```

```
[1] 0.7763719
```

```
cor.test(`OVO (m3)`,Oxig)
```

Pearson's product-moment correlation

data: OVO (m3) and Oxig

t = 7.5933, df = 38, p-value = 3.93e-09

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.6131515 0.8760187

sample estimates:

cor

0.7763719

```
regressao1=lm(Oxig~`OVO (m3)`, data=peixes1)
```

```
regressao1
```

Call:

```
lm(formula = Oxig ~ `OVO (m3)`, data = peixes1)
```

Coefficients:

```
(Intercept)  `OVO (m3)`  
       7.789      0.274
```

```
summary(regressao1)
```

Call:

```
lm(formula = Oxig ~ `OVO (m3)`, data = peixes1)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.0477	-0.5125	-0.0742	0.5419	1.4448

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.78920	0.13375	58.238	< 2e-16 ***
`OVO (m3)`	0.27401	0.03609	7.593	3.93e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6532 on 38 degrees of freedom

Multiple R-squared: 0.6028, Adjusted R-squared: 0.5923

F-statistic: 57.66 on 1 and 38 DF, p-value: 3.93e-09

2. Baixe no site https://www.openintro.org/stat/data/?data=gun_violence_us a base de dados gun_violence_us.csv (GUN VIOLENCE IN THE UNITED STATES), apresentando o relacionamento entre propriedade de armas e violência nos Estados Unidos. A variável dependente é a taxa de propriedade de armas (ownership_rate: percentual de adultos em cada estado que é proprietário de uma arma em 2013) e a variável independente é a taxa de mortalidade (mortality_rate: número de mortes por 100.000 em cada estado em 2014).

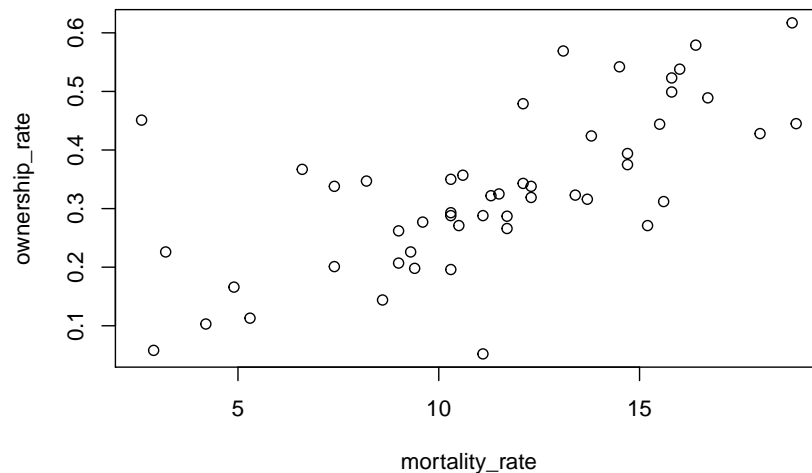
```
library(readr)
gunviolence <- read_csv("https://www.openintro.org/stat/data/gun_violence_us.csv")
head(gunviolence)
```

```
# A tibble: 6 x 3
  state mortality_rate ownership_rate
  <chr>          <dbl>          <dbl>
1 AL             16.7             0.489
2 AK             18.8             0.617
3 AZ             13.4             0.323
4 AR             16.4             0.579
5 CA              7.4             0.201
6 CO             12.1             0.343
```

```
attach(gunviolence)
```

2.1 Faça um diagrama de dispersão para visualizar a relação da taxa e ano. O que você pode concluir?

```
plot(mortality_rate, ownership_rate)
```



2.2 Calcule o coeficiente de correlação linear. Conclua sobre ele.

```
cor(ownership_rate, mortality_rate)
```

```
[1] 0.678234
```

```
cor.test(ownership_rate, mortality_rate)
```

Pearson's product-moment correlation

```
data: ownership_rate and mortality_rate
t = 6.3945, df = 48, p-value = 6.243e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4929473 0.8046719
sample estimates:
```

```
cor
0.678234
```

2.3 Encontre e interprete a equação ajustada.

```
regressaolinear = lm(ownership_rate ~ mortality_rate, data = gunviolence)
regressaolinear
```

Call:

```
lm(formula = ownership_rate ~ mortality_rate, data = gunviolence)
```

Coefficients:

```
(Intercept) mortality_rate
0.07605      0.02245
```

2.4 Apresente e interprete o coeficiente de determinação (R^2).

```
summary(regressaolinear)
```

Call:

```
lm(formula = ownership_rate ~ mortality_rate, data = gunviolence)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-0.27322 -0.05794 -0.01531  0.06196  0.31659
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.07605     0.04232   1.797  0.0786 .
mortality_rate 0.02245     0.00351   6.394 6.24e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.1005 on 48 degrees of freedom

Multiple R-squared: 0.46, Adjusted R-squared: 0.4488

F-statistic: 40.89 on 1 and 48 DF, p-value: 6.243e-08

2.5 Teste a significância da equação de regressão através da ANOVA.

```
anova(regressaolinear)
```

Analysis of Variance Table

Response: ownership_rate

```
      Df Sum Sq Mean Sq F value    Pr(>F)
mortality_rate 1 0.41291 0.41291 40.889 6.243e-08 ***
Residuals    48 0.48472 0.01010
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.6 Faça o intervalo de predição para $x = 0.50$.

```
x = data.frame(mortality_rate = 50)

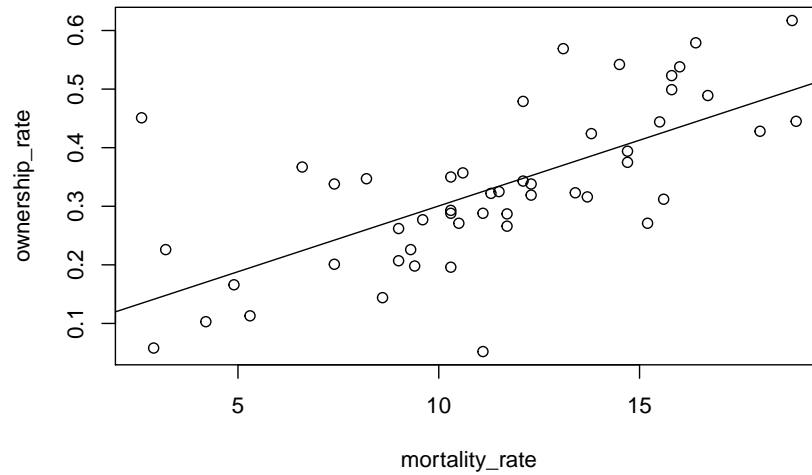
predict(regressaolinear, x, interval = "prediction")
```

```
      fit      lwr      upr
```

```
1 1.198428 0.8577727 1.539084
```

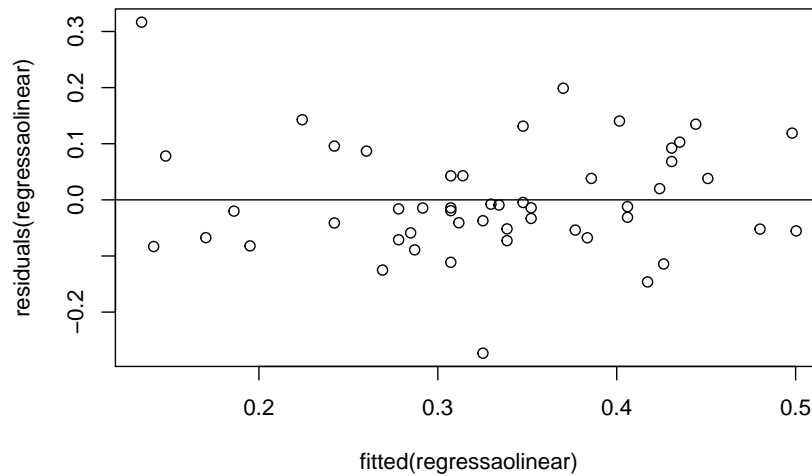
2.7 Trace a reta de regressão ajustada no diagrama de dispersão.

```
plot(mortality_rate,ownership_rate)
abline(regressaolinear)
```

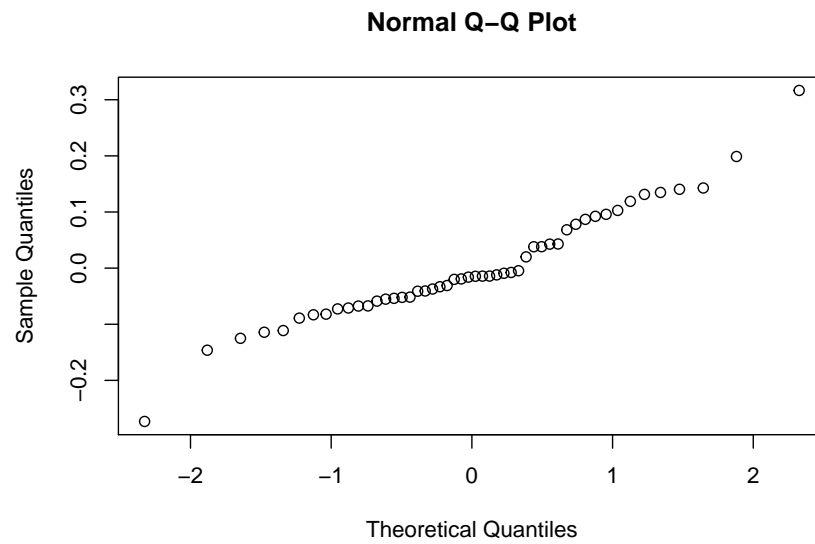


2.8 Faça análise de resíduos.

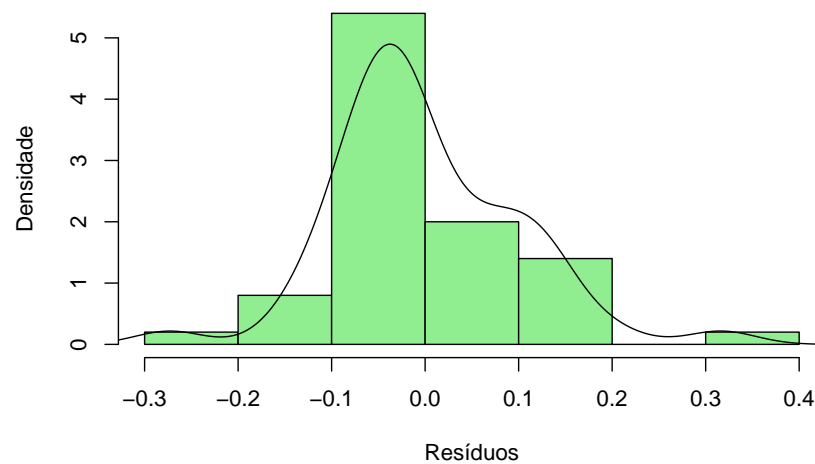
```
plot(fitted(regressaolinear), residuals(regressaolinear))
abline(h = 0)
```



```
qqnorm(residuals(regressaolinear))
```



```
hist(x = regressaolinear$residuals,
     xlab = "Resíduos",
     ylab = "Densidade",
     main = "",
     col = "lightgreen",
     probability = TRUE)
lines(density(regressaolinear$residuals))
```



```
shapiro.test(residuals(regressaolinear))
```

Shapiro-Wilk normality test

```
data: residuals(regressaolinear)
W = 0.95529, p-value = 0.05649
```

```
rstudent(regressaolinear)
```

1	2	3	4	5	6
0.38627088	1.24622376	-0.53867770	1.39047057	-0.41423950	-0.04642843
7	8	9	10	11	12
-0.20484427	-2.96036177	-0.09149499	-0.67783984	3.78123096	2.07050632
13	14	15	16	17	18
-0.16053053	-0.14091135	0.97234523	-0.07667445	0.38172103	-0.57275489
19	20	21	22	23	24
-0.58870945	-0.71329662	0.81708490	-0.37076516	1.47371549	-0.53479043
25	26	27	28	29	30
-1.50324682	0.93805726	-0.89550830	-0.31114358	-1.27160291	-0.84142225
31	32	33	34	35	36
0.69116269	-0.69570477	-0.51561770	1.33123843	-1.12220720	-1.16573372
37	38	39	40	41	42
-0.72712288	-0.40623214	-0.87353808	0.20131513	0.42624624	-0.12051860
43	44	45	46	47	48
0.42876123	-0.33036219	-0.19178895	-0.14197513	-0.14500881	1.43636497
49	50				
0.87667170	1.04851134				

```
plot(rstandard(regressaolinear))
abline(h=2, col="red")
abline(h=-2, col="red")
```

