Regressão Linear Simples

Cristian Villegas

2022-09-26

```
regressao<- function(x, y){
n<- length(x)
Sxy<- sum(x*y)-n*mean(x)*mean(y)
Sxx<- sum(x*x)-n*mean(x)*mean(x)
b<- Sxy/Sxx
a<- mean(y)-b*mean(x)
saidas<- list()
saidas$Sxy<- Sxy
saidas$Sxx<- Sxx
saidas$mediaX<- mean(x)
saidas$mediaY<- mean(y)
saidas$n<- length(x)
saidas$a<- a
saidas$b<- b
return(saidas)
}

correlacao<- function(x, y){
n<- length(x)
Sxy<- sum(x*y)-n*mean(x)*mean(y)
Sxx<- sum(x*x)-n*mean(x)*mean(x)
Syy<- sum(y*y)-n*mean(y)*mean(y)
r<- Sxy/sqrt(Sxx*Syy)

saidas<- list()
saidas$Sxy<- Sxy
saidas$Sxx<- Sxx
saidas$Syy<- Syy
saidas$mediaX<- mean(x)
saidas$mediaY<- mean(y)
saidas$n<- length(x)
saidas$r<- r
return(saidas)
}
```

```
dados<- cars[1:10,]
dados
```

```
##    speed dist
## 1      4    2
## 2      4   10
```

```
## 3      7    4
## 4      7   22
## 5      8   16
## 6      9   10
## 7     10   18
## 8     10   26
## 9     10   34
## 10    11   17
```

```
(soma_dados<- with(dados,
     data.frame(x=speed,
           y=dist,
           xy=speed*dist,
           x2=speed^2,
           y2=dist^2)))
```

```
##     x  y  xy  x2   y2
## 1   4  2   8  16    4
## 2   4 10  40  16  100
## 3   7  4  28  49   16
## 4   7 22 154  49  484
## 5   8 16 128  64  256
## 6   9 10  90  81  100
## 7  10 18 180 100  324
## 8  10 26 260 100  676
## 9  10 34 340 100 1156
## 10 11 17 187 121  289
```

```
apply(soma_dados, 2, sum)
```

```
##    x    y   xy   x2   y2
##   80  159 1415  696 3405
```

```
regressao(dados$speed, dados$dist)
```

```
## $Sxy
## [1] 143
##
## $Sxx
## [1] 56
##
## $mediaX
## [1] 8
##
## $mediaY
## [1] 15.9
##
## $n
## [1] 10
##
## $a
## [1] -4.528571
```

```
##
## $b
## [1] 2.553571
```

```
correlacao(dados$speed, dados$dist)
```

```
## $Sxy
## [1] 143
##
## $Sxx
## [1] 56
##
## $Syy
## [1] 876.9
##
## $mediaX
## [1] 8
##
## $mediaY
## [1] 15.9
##
## $n
## [1] 10
##
## $r
## [1] 0.6453079
```

```
summary(dados)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   : 2.0
##  1st Qu.: 7.0   1st Qu.:10.0
##  Median : 8.5   Median :16.5
##  Mean   : 8.0   Mean   :15.9
##  3rd Qu.:10.0   3rd Qu.:21.0
##  Max.   :11.0   Max.   :34.0
```

```
cor(dados$speed, dados$dist)
```

```
## [1] 0.6453079
```

```
ajuste<- lm(dist~speed, data= dados)
coef(ajuste)
```

```
## (Intercept)       speed
##   -4.528571    2.553571
```

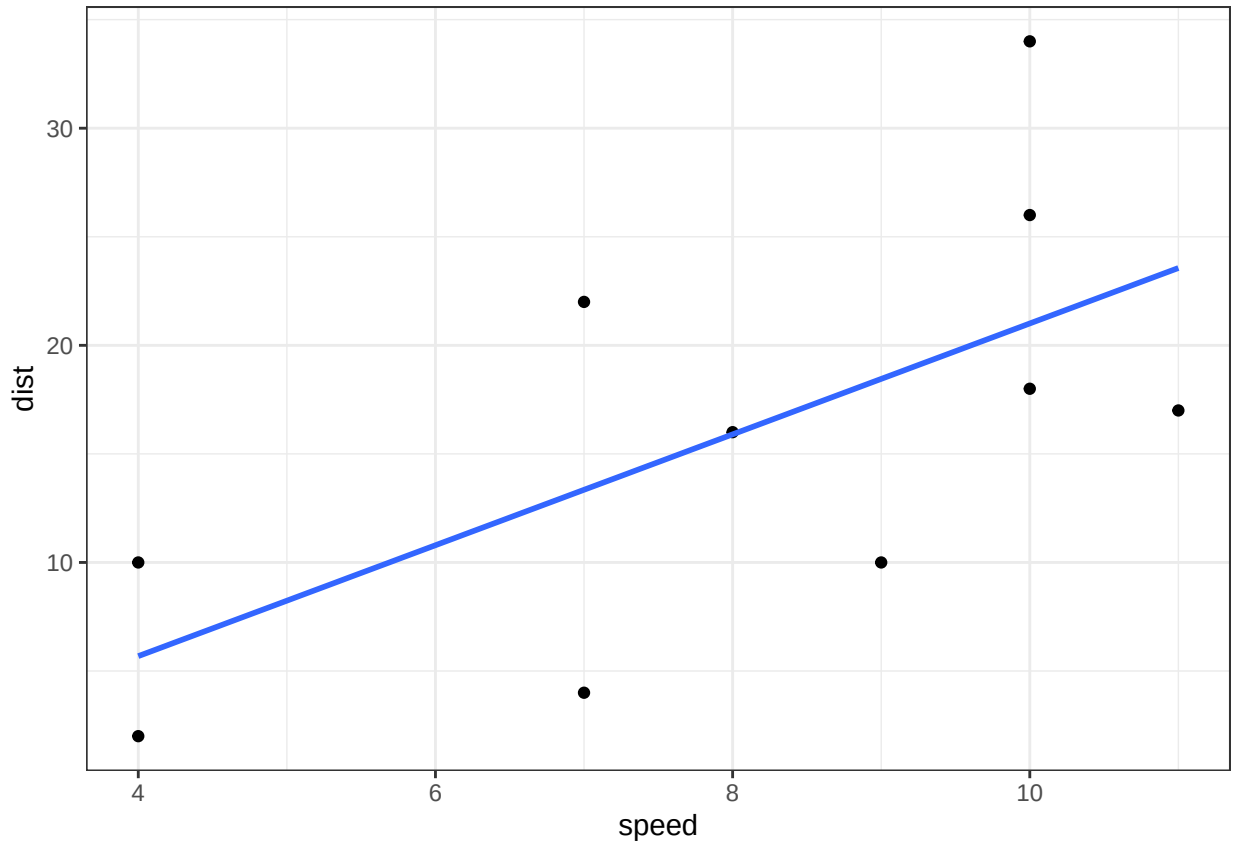
```
summary(ajuste)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = dados)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.346 -5.842 -1.454  4.823 12.993
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.529      8.916  -0.508   0.6252
## speed          2.554      1.069   2.389   0.0439 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.998 on 8 degrees of freedom
## Multiple R-squared:  0.4164, Adjusted R-squared:  0.3435
## F-statistic: 5.709 on 1 and 8 DF,  p-value: 0.04391
```

```
predict(ajuste,
        newdata = data.frame(speed=c(9,10)),
        data= dados)
```

```
##        1        2
## 18.45357 21.00714
```

```
library(ggplot2)
ggplot(dados, aes(speed, dist))+
  geom_point()+
  geom_smooth(method = "lm", se=FALSE)+
  theme_bw()
```


dist
30
20
10
4
6
8
10
speed