

Regressão e correlação

Cristian Villegas e Gustavo Jun Yakushiji

19/Setembro/2022

Sumário

1	Regressão linear simples	1
2	Correlação linear de Pearson	4

1 Regressão linear simples

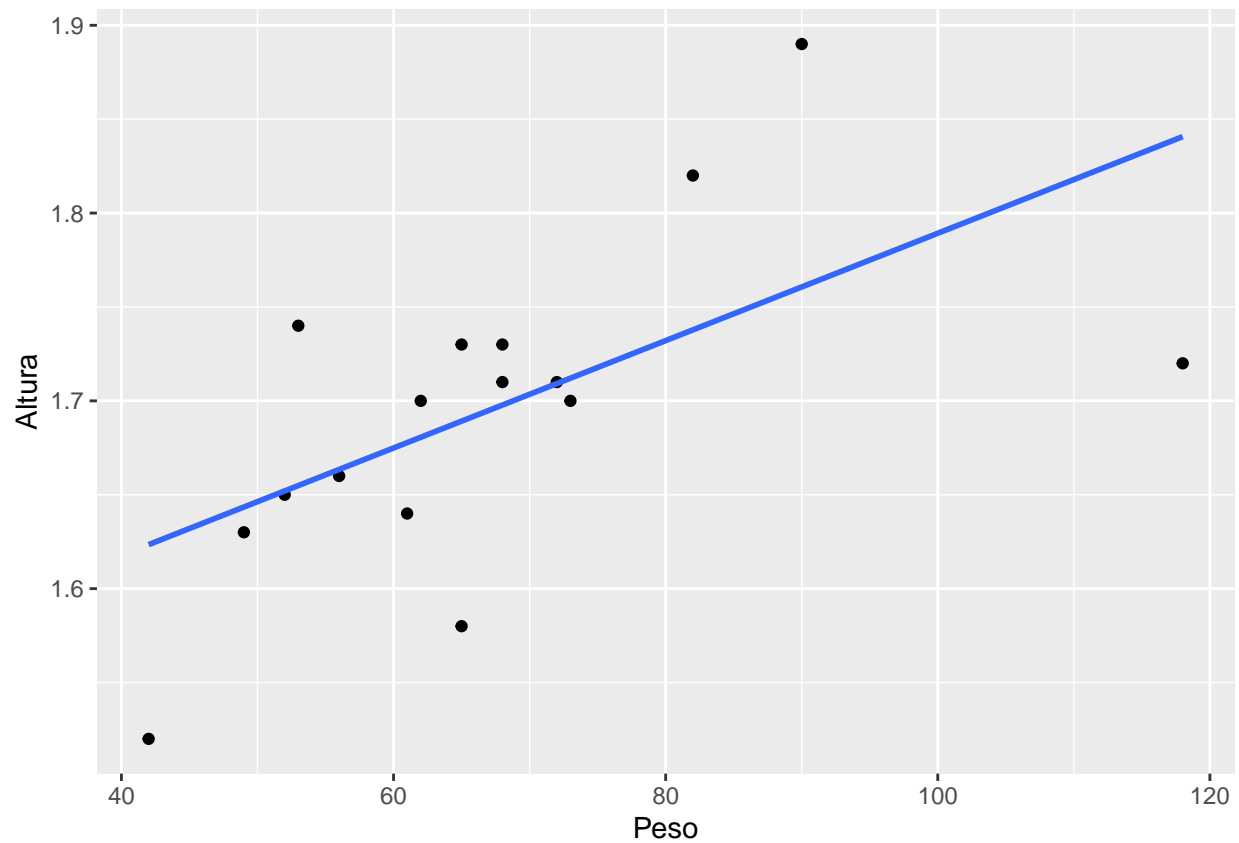
O código a seguir foi criado pelo *Gustavo Jun Yakushiji* com a supervisão do *Cristian Villegas*

```
#1. Importação de dados
library(readxl)
(dados_alunos <- read_excel("dados_alunos.xlsx"))
```

```
## # A tibble: 16 x 2
##   Peso Altura
##   <dbl> <dbl>
## 1    68   1.73
## 2    42   1.52
## 3    82   1.82
## 4    49   1.63
## 5    65   1.58
## 6   118   1.72
## 7    72   1.71
## 8    68   1.71
## 9    65   1.73
## 10   73   1.7
## 11   56   1.66
## 12   90   1.89
## 13   62   1.7
## 14   61   1.64
## 15   53   1.74
## 16   52   1.65
```

```
# 2. Gráfico de dispersão + regressão
library(ggplot2)
ggplot(data = dados_alunos,
       aes(x = Peso, y = Altura))+
```

```
geom_point()+
geom_smooth(method = "lm", se = FALSE)
```



```
#3. Calcular regressão na unha
##Calcular média de x=Peso e y=Altura
(media_x <- mean(dados_alunos$Peso))
```

```
## [1] 67.25
```

```
(media_y <- mean(dados_alunos$Altura))
```

```
## [1] 1.695625
```

```
## Calcular média de Peso ao quadrado
mean(dados_alunos$Peso)^2
```

```
## [1] 4522.562
```

```
## Calcular média de Altura ao quadrado
mean(dados_alunos$Altura)^2
```

```
## [1] 2.875144
```

```
# Banco de dados com x, y e x*y
(dados <- data.frame(
  x = dados_alunos$Peso,
  y = dados_alunos$Altura,
  x.y = dados_alunos$Peso*dados_alunos$Altura,
  x_quadrado = dados_alunos$Peso^2,
  y_quadrado = dados_alunos$Altura^2
))
```

```
##      x      y      x.y x_quadrado y_quadrado
## 1  68 1.73 117.64      4624      2.9929
## 2  42 1.52  63.84      1764      2.3104
## 3  82 1.82 149.24      6724      3.3124
## 4  49 1.63  79.87      2401      2.6569
## 5  65 1.58 102.70      4225      2.4964
## 6 118 1.72 202.96     13924      2.9584
## 7  72 1.71 123.12      5184      2.9241
## 8  68 1.71 116.28      4624      2.9241
## 9  65 1.73 112.45      4225      2.9929
## 10 73 1.70 124.10      5329      2.8900
## 11 56 1.66  92.96      3136      2.7556
## 12 90 1.89 170.10      8100      3.5721
## 13 62 1.70 105.40      3844      2.8900
## 14 61 1.64 100.04      3721      2.6896
## 15 53 1.74  92.22      2809      3.0276
## 16 52 1.65  85.80      2704      2.7225
```

```
# Feito pelo Cristian após a aula
apply(dados, 2, sum)
```

```
##      x      y      x.y x_quadrado y_quadrado
## 1076.0000  27.1300 1838.7200 77338.0000  46.1159
```

```
# Soma de x*y
(soma_xy <- sum(dados$x.y))
```

```
## [1] 1838.72
```

```
# Soma x_quadrado
(soma_xquad <- sum(dados$x_quadrado))
```

```
## [1] 77338
```

```
# Calculo b
(b <- (soma_xy-nrow(dados_alunos)*media_x*media_y)/(soma_xquad-nrow(dados_alunos)*media_x^2))
```

```
## [1] 0.00285865
```

```
# Calculo a
(a <- media_y - b*media_x)
```

```
## [1] 1.503381
```

```
# Equação:  $y = a + b \cdot x$ 
(peso80 <- a + b*80)
```

```
## [1] 1.732073
```

```
(peso100 <- a + b*100)
```

```
## [1] 1.789246
```

```
# Função  $Altura = a + b \cdot Peso$ 
lm(dados_alunos$Altura~dados_alunos$Peso)
```

```
##
## Call:
## lm(formula = dados_alunos$Altura ~ dados_alunos$Peso)
##
## Coefficients:
##      (Intercept)  dados_alunos$Peso
##      1.503381      0.002859
```

```
# Calculando correlação
```

```
(soma_yquad <- sum(dados$y_quadrado))
```

```
## [1] 46.1159
```

```
(syy <- soma_yquad-nrow(dados_alunos)*media_y^2)
```

```
## [1] 0.1135937
```

```
(sxx <- soma_xquad-nrow(dados_alunos)*media_x^2)
```

```
## [1] 4977
```

```
(sxy <- soma_xy-nrow(dados_alunos)*media_x*media_y)
```

```
## [1] 14.2275
```

2 Correlação linear de Pearson

```
(correlacao <- sxy/sqrt(sxx*syy))
```

```
## [1] 0.5983669
```

```
cor(dados_alunos$Altura, dados_alunos$Peso)
```

```
## [1] 0.5983669
```