

Teste de permutações para associação

2020

Analisamos os resultados da prova de patinação artística individual feminina na Olimpíada de Inverno de Sochi, 2014. Os dados estão no arquivo <https://arxiv.org/pdf/1411.5279.pdf>. Utilizamos simulação de Monte Carlo implementada em linguagem R.

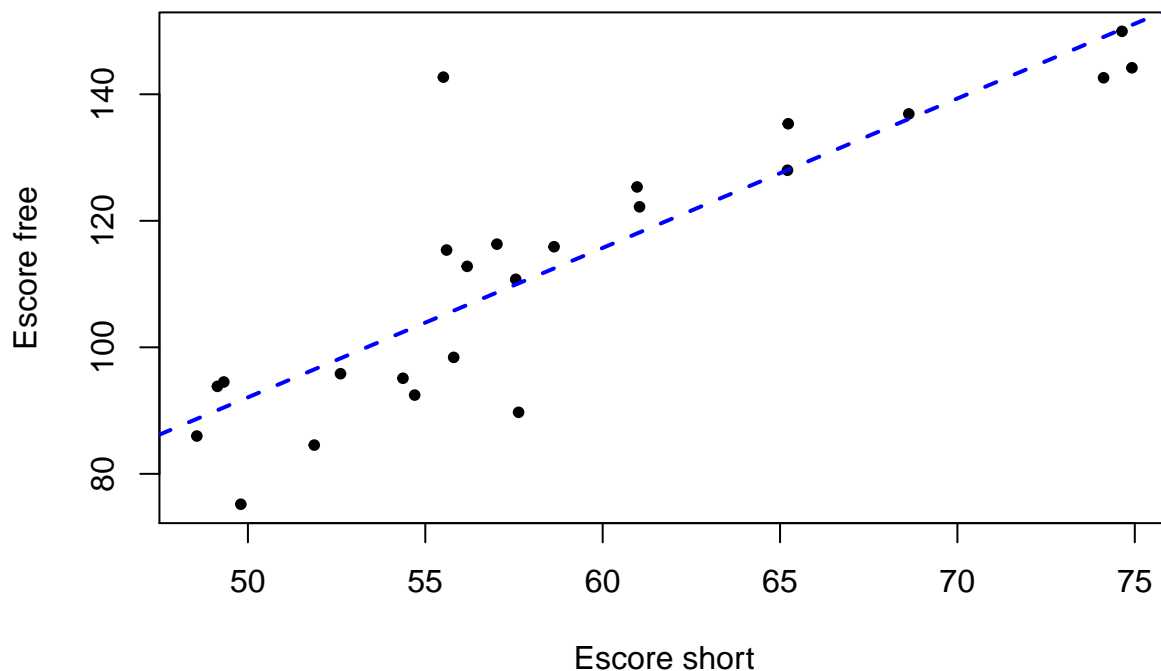
```
# Separador decimal: ","  
options(OutDec = ",")
```

A pontuação de cada atleta é a soma dos escores *short* e *free*.

```
short <- c(  
74.64, 74.92, 74.12, 68.63, 65.23, 55.51, 65.21, 60.97, 61.04,  
58.63, 57.02, 55.60, 56.18, 57.55, 55.80, 54.37, 52.61, 57.63,  
54.70, 49.32, 49.14, 51.87, 48.56, 49.80)  
  
free <- c(  
149.95, 144.19, 142.61, 136.90, 135.34, 142.71, 127.99, 125.35,  
122.21, 115.90, 116.31, 115.38, 112.80, 110.75, 98.41, 95.11,  
95.83, 89.73, 92.45, 94.52, 93.83, 84.55, 85.98, 75.20)
```

O gráfico de dispersão permite uma primeira avaliação sobre a associação entre as duas variáveis. A linha tracejada representa o ajuste pelo método dos mínimos quadrados.

```
plot(short, free, pch = 20, xlab = "Escore short", ylab = "Escore free")  
abline(lm(free ~ short), col = "blue", lty = 2, lwd = 2)
```



Com base no gráfico acima, percebemos que o coeficiente de correlação de Pearson é adequado como medida

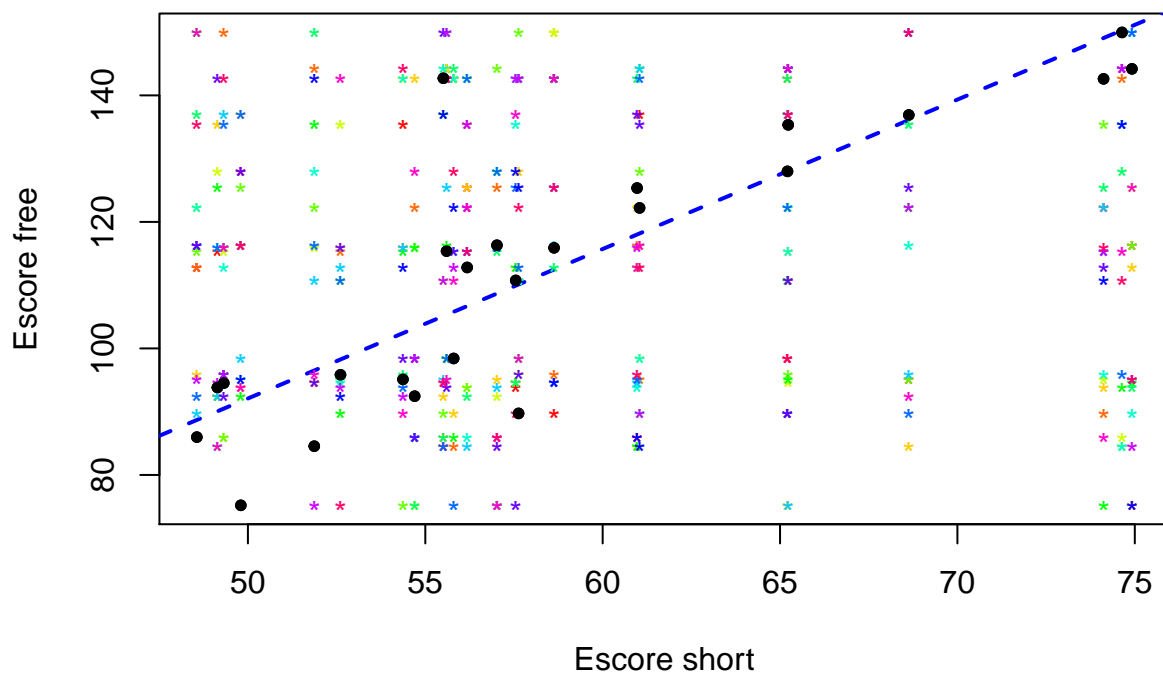
de associação.

```
n <- length(free)
dobs <- cor(short, free, method = "pearson")
cat("\n n =", n,
    "\n coeficiente de correlação de Pearson:", round(dobs, 3))
```

```
##
## n = 24
## coeficiente de correlação de Pearson: 0,858
```

O gráfico abaixo inclui as 15 primeiras das $R = 9999$ permutações utilizadas.

```
plot(short, free, pch = 20, xlab = "Escore short", ylab = "Escore free")
abline(lm(free ~ short), col = "blue", lty = 2, lwd = 2)
set.seed(1809)
R <- 9999
mycolors <- rainbow(15)
D <- c()
for (m in 1:R) {
  short1 <- sample(short)
  if (m <= 15) {
    points(short1, free, pch = "*", col = mycolors[m])
  }
  D[m] <- cor(short1, free, method = "pearson")
}
points(short, free, pch = 20)
```



```
summary(D)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -0,7402151 -0,1496233 -0,0031529 -0,0009908  0,1464842  0,6948482
```

Em seguida calculamos uma aproximação do valor- p do teste para a hipótese alternativa bilateral.

```

k1 <- sum(D >= abs(dobs))
k2 <- sum(D <= -abs(dobs))
k <- min(k1, k2)
valp <- 2 * (k + 1) / (R + 1)
cat("\n p (bilateral):", valp)

```

```

##
## p (bilateral): 2e-04

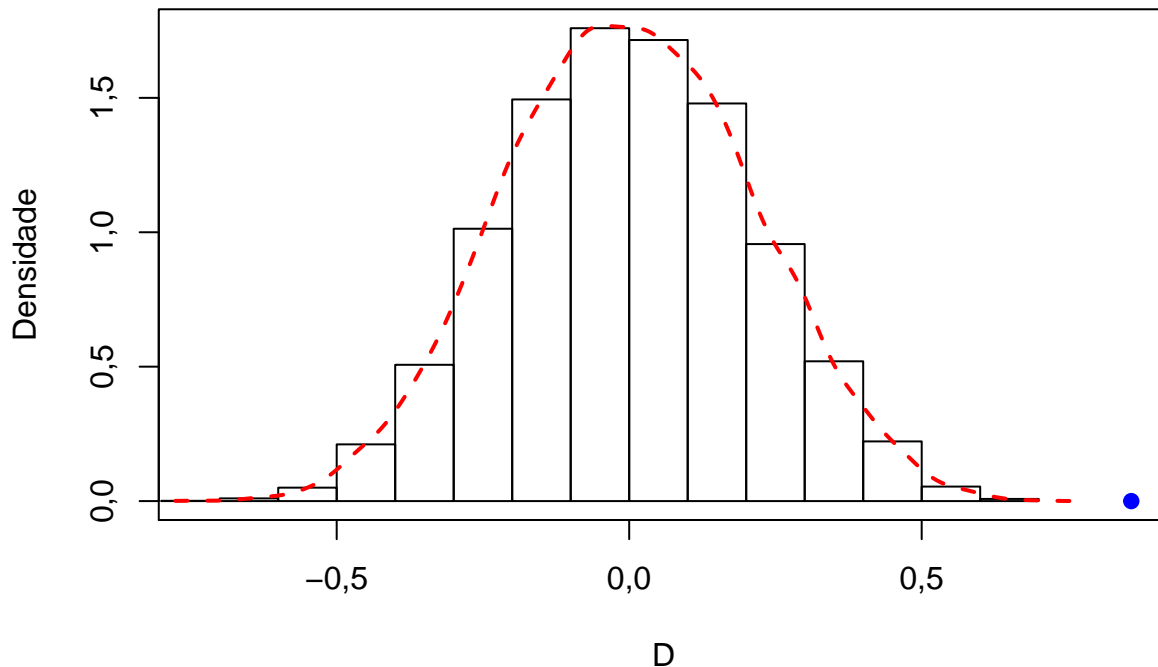
```

Finalmente apresentamos o histograma da estatística de teste (D), uma estimativa de sua função densidade e o valor observado.

```

hist(D, main = "", freq = FALSE, ylab = "Densidade",
      xlim = range(dobs, D))
lines(density(D), lty = 2, col = "red", lwd = 2)
points(dobs, 0, pch = 19, col = "blue")
box()

```



Nota 1. Com base nos dois gráficos anteriores, o resultado do teste é surpresa?

Nota 2. Refaça o exemplo com $R = 999$ permutações de Monte Carlo e compare os resultados.

Nota 3 Refaça o exemplo utilizando `free <- rnorm(n, mean = 100, sd = 15)`. Os resultados estão de acordo com o que você esperava?