Teste de permutações para associação

2023

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.

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

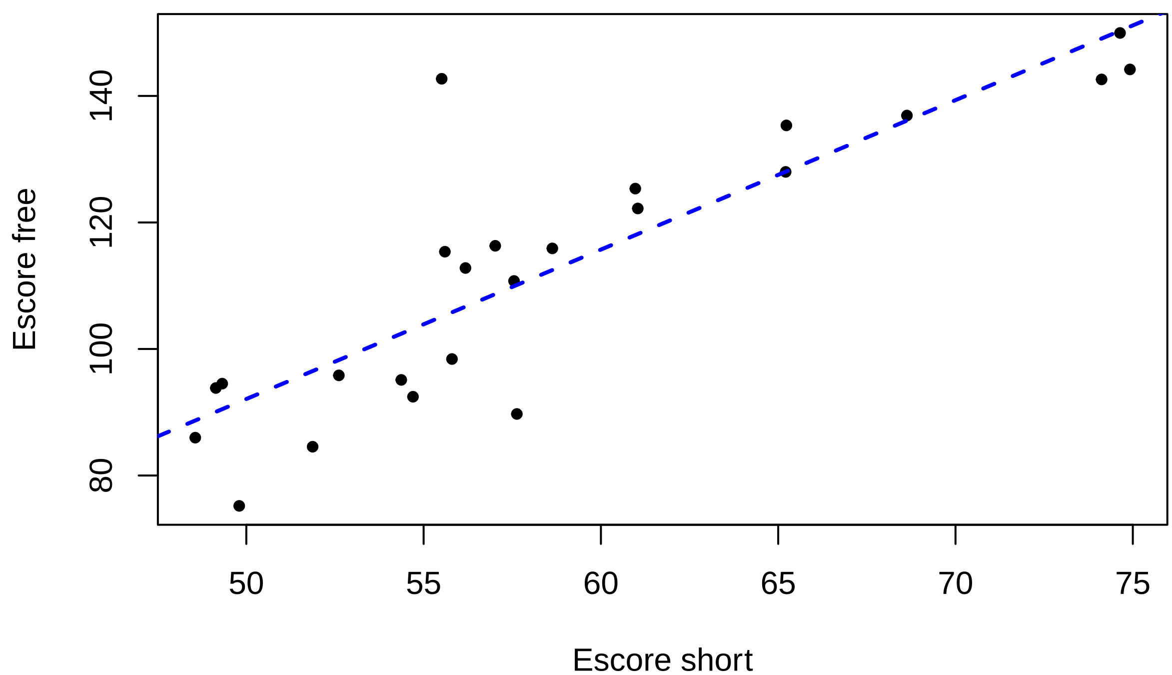
A pontuação de cada atleta é a soma dos escores *short* e *free*.

```r
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.

```r
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 $\rho = Cov(X, Y)/\sqrt{Var(X)Var(Y)}$

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 número de permutações é 24!, que é da ordem de 10^{23}.

```
cat("\n Número de permutações =", factorial(n))
```

```
##
##  Número de permutações = 6,204484e+23
```

A distribuição do coeficiente de correlação de Pearson amostral $\widehat{\rho}$ é aproximada utilizando $R = 9999$ permutações.

```
set.seed(76170)
R <- 9999
D <- c()
for (m in 1:R) {
   freel <- sample(free)
   D[m] <- cor(short, freel, method = "pearson")
}
summary(D)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0,723882 -0,144057 -0,002037 -0,002069  0,136830  0,692018
```

Em seguida calculamos uma aproximação do valor-p do teste para a hipótese alternativa bilateral ($\mathrm{H}_0 : \rho = 0$ contra $\mathrm{H}_1 : \rho \neq 0$) considerando $D = \widehat{\rho}$ como estatística de teste.

```
k1 <- sum(D >= abs(dobs))
k2 <- sum(D <= -abs(dobs))
cat("\n k1, k2 =", c(k1, k2))
```

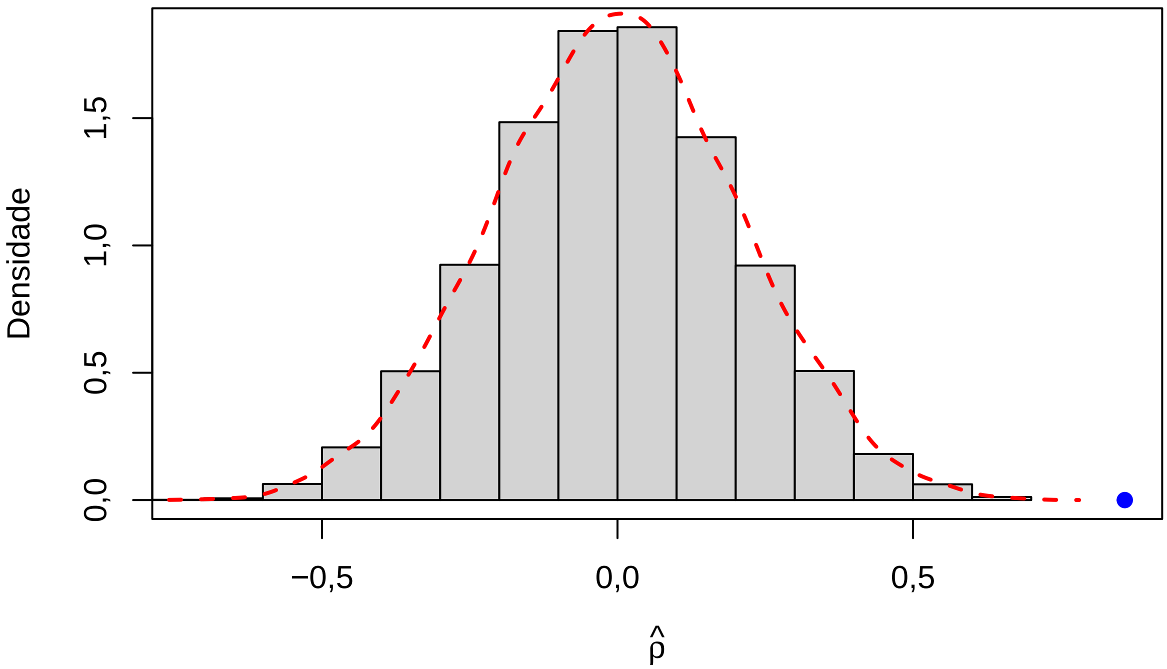
```
##
##  k1, k2 = 0 0
```

```
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 obeservado.

```
hist(D, main = "", freq = FALSE, ylab = "Densidade",
     xlim = range(dobs, D), xlab = expression(hat(rho)))
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 utilizando `free <- rnorm(n, mean = 100, sd = 15)`. Os resultados estão de acordo com o que você esperava?