

Coeficiente de correlação de concordância e *bootstrap*

## Concordância

Uma forma de quantificar a **concordância** entre medições de variáveis  $X$  e  $Y$  é dada pelo **coeficiente de correlação de concordância**, definido como

## Concordância

Uma forma de quantificar a **concordância** entre medições de variáveis  $X$  e  $Y$  é dada pelo **coeficiente de correlação de concordância**, definido como

$$\rho_c = 1 - \frac{\text{DQM}}{\text{DQM}_0} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2}, \quad (1)$$

em que DQM denota o desvio quadrático médio

$$\begin{aligned} \text{DQM} &= E\{(X - Y)^2\} \\ &= \{E(X - Y)\}^2 + \text{var}(X - Y) \\ &= (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} \end{aligned}$$

e  $\text{DQM}_0$  significa que a esperança é calculada supondo que a covariância entre  $X$  e  $Y$  é nula ( $\sigma_{XY} = 0$ ).

## Concordância

Uma forma de quantificar a **concordância** entre medições de variáveis  $X$  e  $Y$  é dada pelo **coeficiente de correlação de concordância**, definido como

$$\rho_c = 1 - \frac{\text{DQM}}{\text{DQM}_0} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2}, \quad (1)$$

em que DQM denota o desvio quadrático médio

$$\begin{aligned} \text{DQM} &= E\{(X - Y)^2\} \\ &= \{E(X - Y)\}^2 + \text{var}(X - Y) \\ &= (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY} \end{aligned}$$

e  $\text{DQM}_0$  significa que a esperança é calculada supondo que a covariância entre  $X$  e  $Y$  é nula ( $\sigma_{XY} = 0$ ).

Pode ser provado que  $\rho_c \in [-1, 1]$ . Os valores extremos representam discordância perfeita e concordância perfeita, respectivamente, enquanto que  $\rho_c = 0$  representa ausência de associação.

$\rho_c = 0$  se, e somente se,  $\rho = 0$  (Pearson), ou seja,  $\sigma_{XY} = 0$ .

$\rho_c = 0$  se, e somente se,  $\rho = 0$  (Pearson), ou seja,  $\sigma_{XY} = 0$ .

$\rho_c = \rho$  se, e somente se,  $\mu_X = \mu_Y$  e  $\sigma_X = \sigma_Y$ .

$\rho_c = 0$  se, e somente se,  $\rho = 0$  (Pearson), ou seja,  $\sigma_{XY} = 0$ .

$\rho_c = \rho$  se, e somente se,  $\mu_X = \mu_Y$  e  $\sigma_X = \sigma_Y$ .

Um **estimador consistente** para  $\rho_c$  é dado pela sua versão amostral, ou seja,

$$r_c = \frac{2S_{XY}}{(\bar{X} - \bar{Y})^2 + S_X^2 + S_Y^2},$$

em que  $S_{XY} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})/n$ ,  $S_X^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$  e

$S_Y^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/n$ .

# Concordância e correlação de Pearson

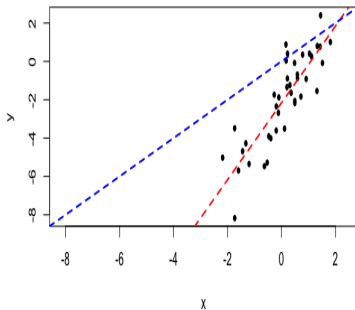


Figura 1: Gráfico de dispersão.

## Concordância e correlação de Pearson

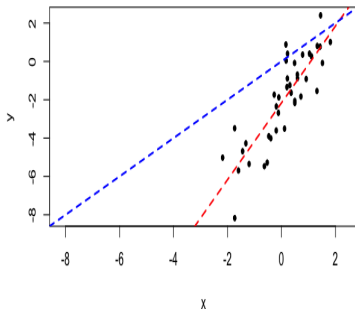


Figura 1: Gráfico de dispersão.

Neste exemplo,  $r = 0,850$  (Pearson) e  $r_c = 0,365$ .

## Problema

Os dados correspondem a **notas de duas provas**  $(X, Y)$  aplicadas a uma turma no ano de 2025.

## Problema

Os dados correspondem a **notas de duas provas** ( $X, Y$ ) aplicadas a uma turma no ano de 2025.

Represente graficamente os dados, apresente estimativas do coeficiente de correlação de concordância e do coeficiente de correlação de Pearson e de seus erros padrão.

```
## Dados
```

```
x <- scan() # Prova 1
```

```
6.50 7.50 8.50 9.00 6.50 6.50 9.00 9.15 2.00 6.50 6.50 9.30 6.50 6.00 8.50
7.00 9.00 6.50 8.50 8.90 7.50 9.00 9.75 9.40 8.90 9.50 9.00 4.45 7.00 6.00
8.50 7.25 6.50 9.55 8.65 7.75 0.50 5.90 9.00 7.50 6.50 6.90 5.00 5.75 2.00
8.50 8.90 8.80 6.40 2.00 6.40 2.50 2.00 6.40
```

```
y <- scan() # Prova 2
```

```
3.50 5.60 6.20 8.60 6.30 4.40 9.40 9.60 0.80 10.00 6.30 9.70 5.50
6.90 9.70 8.00 7.70 2.50 7.10 6.70 4.85 7.50 1.20 9.00 5.90 4.00
10.00 8.90 3.00 0.30 5.60 7.90 6.10 9.50 7.20 6.25 5.50 2.30 6.60
6.20 4.90 4.60 5.10 7.20 2.30 4.90 6.00 7.50 7.40 4.90 2.90 9.40
5.90 4.20
```

```
cat("\n n =", n <- length(x), "\n Média (prova 1):", mean(x),
    "\n Média (prova 2):", mean(y),
    "\n Desvio padrão (prova 1):", sd(x),
    "\n Desvio padrão (prova 2):", sd(y),
    "\n r (Pearson):", cor(x, y))
```

$n = 54$

Média (prova 1): 6.990741

Média (prova 2): 6.101852

Desvio padrão (prova 1): 2.259217

Desvio padrão (prova 2): 2.456354

$r$  (Pearson): 0.3318317

```
## Gráfico de dispersão
rxy <- range(x, y)
plot(x, y, pch = 20, xlim = rxy, ylim = rxy,
      xlab = "Prova 1", ylab = "Prova 2")
abline(0, 1, lty = 2, lwd = 2, col = "blue")
abline(lm(y ~ x), lty = 2, lwd = 2, col = "red")
```

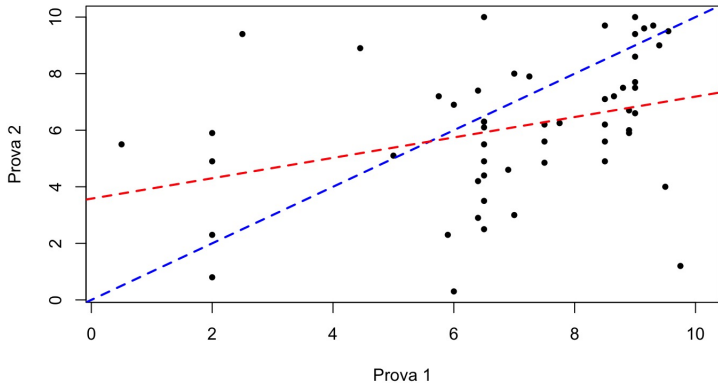


Figura 2: Gráfico de dispersão das notas de duas provas (—: concordância e —: modelo linear).

```
## Estimativa do coef. de correlação de concordância
rc <- function(X, Y) {
  n <- length(X)
  cn <- (n - 1) / n
  Sx2 <- var(X) * cn
  Sy2 <- var(Y) * cn
  Sxy <- cov(X, Y) * cn
  return(2 * Sxy / ((mean(X) - mean(Y))^2 + Sx2 + Sy2))
}
```

```
## Bootstrap
set.seed(77)
B <- 5000 # Número de amostras
rs <- rcs <- c() # Estimativas bootstrap
for (b in 1:B) {
  ib <- sample(n, n, replace = TRUE)
  xs <- x[ib] # x*
  ys <- y[ib] # y*
  rs[b] <- cor(xs, ys)
  rcs[b] <- rc(xs, ys)
}
```

```
## Histogramas
rx <- range(rs, rcs)
hist(rs, freq = FALSE, main = "", xlab = expression(r*"),
     ylab = "Densidade", col = "lightgreen", xlim = rx)
lines(density(rs), col = "blue", lty = 2, lwd = 2)
box()
hist(rcs, freq = FALSE, main = "", xlab = expression(r[c]*"),
     ylab = "Densidade", col = "lightyellow", xlim = rx)
lines(density(rcs), col = "blue", lty = 2, lwd = 2)
box()
```

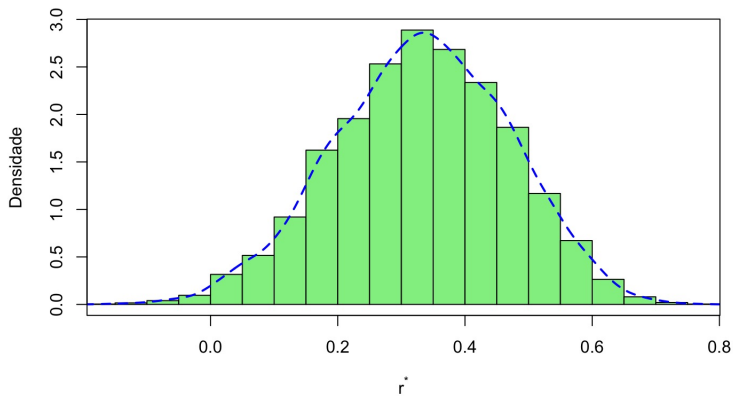


Figura 3: Distribuição das estimativas *bootstrap*  $r^*$ .

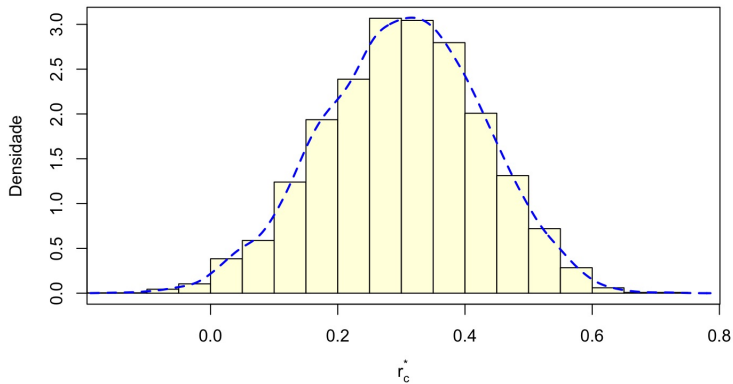


Figura 4: Distribuição das estimativas *bootstrap*  $r_c^*$ .

```
## Resultados
cat("\n B:", B, "\n")
cat("\n Estimativa (Pearson) :", cor(x, y), "\n e.p. bootstrap:",
    sd(rs))
cat("\n Estimativa (concordancia):", rc(x, y), "\n e.p. bootstrap:",
    sd(rcs))
```

```
## Resultados
cat("\n B:", B, "\n")
cat("\n Estimativa (Pearson) :", cor(x, y), "\n e.p. bootstrap:",
    sd(rs))
cat("\n Estimativa (concordancia):", rc(x, y), "\n e.p. bootstrap:",
    sd(rcs))
```

```
B: 5000
Estimativa (Pearson) : 0.3318317
e.p. bootstrap: 0.1372994
Estimativa (concordancia): 0.308384
e.p. bootstrap: 0.1252033
```

- ① A distribuição assintótica de  $r_c$  foi estudada por Lin, Hedayat, Sinha, and Yang (2002, Statistical methods in assessing agreement: models, issues, and tools, *Journal of the American Statistical Association* 97, 257–270).

- 1 A distribuição assintótica de  $r_c$  foi estudada por Lin, Hedayat, Sinha, and Yang (2002, Statistical methods in assessing agreement: models, issues, and tools, *Journal of the American Statistical Association* 97, 257–270).
- 2 A função `boot` do pacote `boot` implementa o método *bootstrap*. Sua forma mais simples é mostrada abaixo.

## Dois coeficientes

```
rcb <- function(mydata, inds) {  
  X <- mydata[inds, 1]  
  Y <- mydata[inds, 2]  
  n <- length(X)  
  cn <- (n - 1) / n  
  Sx2 <- var(X) * cn  
  Sy2 <- var(Y) * cn  
  Sxy <- cov(X, Y) * cn  
  return(c(cor(X, Y),  
           2 * Sxy / ((mean(X) - mean(Y))^2 + Sx2 + Sy2)))  
}
```

```
mydata <- data.frame(cbind(x, y))  
  
set.seed(77)  
boot(mydata, statistic = rcb, R = 5000)
```

## ORDINARY NONPARAMETRIC BOOTSTRAP

Call:

```
boot(data = mydata, statistic = rcb, R = 5000)
```

Bootstrap Statistics :

|     | original  | bias          | std. error |
|-----|-----------|---------------|------------|
| t1* | 0.3318317 | -0.0005857648 | 0.1367010  |
| t2* | 0.3083840 | -0.0073328767 | 0.1256044  |