

Bootstrap

2017

Concordância

Uma forma de quantificar a **concordância entre medições** obtidas por dois métodos diferentes (X e Y) é dada pelo **coeficiente de correlação de concordância**, definido como

$$\rho_c = 1 - \frac{\text{DQM}}{\text{DQM}|_{\sigma_{XY}=0}} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2},$$

em que DQM denota o desvio quadrático médio

$$\text{DQM} = E[(X - Y)^2] = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$$

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

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.

Concordância

Um **estimador** para ρ_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

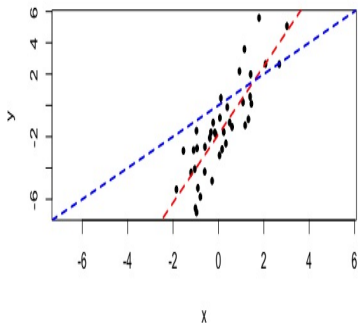


Figura 1: Gráfico de dispersão.

Neste exemplo, $r = 0,944$ (Pearson) e $r_c = 0,450$.

Problema

Foi coletada uma amostra aleatória de pares (X, Y) de **medições angulares** por dois métodos, sendo que o primeiro método (identificado por X) é mais caro do que o segundo método.

Amostra de pares (x, y)

x : 100, 58, 95, 55, 79, 95, 60, 88, 68, 94, 60, 64, 88, 57, 66, 67,
76, 95, 85, 105, 80, 85, 82, 102, 100, 75, 40, 70, 63, 103, 95,
80, 72, 68, 48, 70, 90, 60, 80, 96, 54, 80, 88, 70, 90, 79, 100,
85, 108, 53, 58, 49

y : 97, 77, 74, 59, 79, 85, 78, 78, 68, 96, 74, 64, 76, 60, 78, 71,
67, 103, 95, 78, 70, 80, 78, 102, 102, 77, 45, 60, 50, 94, 91,
66, 63, 65, 58, 75, 105, 65, 80, 90, 58, 75, 83, 78, 85, 65, 90,
76, 100, 65, 40, 53

Represente graficamente os dados, apresente uma estimativa do coeficiente de correlação de concordância e de seu erro padrão.

Solução

```
## Dados
x <- c(100, ..., 49)
y <- c(97, ..., 53)

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

n = 52

Média (método 1): 77.46154

Média (método 2): 75.78846

r (Pearson): 0.8196273

Solução

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

Solução

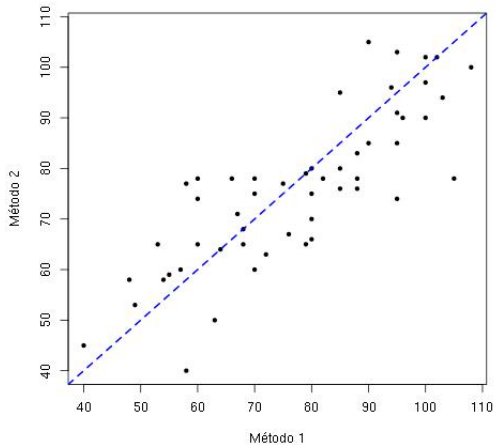


Figura 2: Gráfico de dispersão de medições angulares (em $^{\circ}$).

Solução

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

Solução

```
## Bootstrap
set.seed(7714)
B <- 5000 # Número de amostras
rcs <- c() # Estimativas bootstrap
for (b in 1:B) {
  rcs[b] <- rc(x, y, sample(n, n, replace = TRUE))
}

hist(rcs, freq = FALSE, main = "", xlab = expression(r[c]^{"*"}),
      ylab = "Densidade", col = "lightyellow")
lines(density(rcs), col = "blue", lty = 2, lwd = 2)
box()
```

Solução

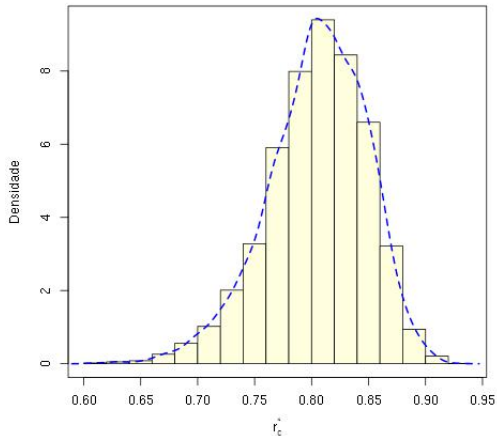


Figura 3: Distribuição das estimativas *bootstrap* r_c^* .

Solução

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

Estimativa: 0.8098709

B: 5000

e.p. bootstrap: 0.04381959

Notas

- 1 Escreva um código para obtenção das estimativas *bootstrap* r_c^* sem utilizar “for” (folha 10).
- 2 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).