

Teste de Kolmogorov-Smirnov em duas amostras

```
### Teste de Kolmogorov-Smirnov (KS) em duas amostras

## Dados
# Problema 34, p. 185 em Hollander & Wolfe (1999, 2nd ed.)
# Concentração de um hormônio (em ng/ml)
# H1 bilateral
tipoA <- c(3.6, 2.6, 4.7, 8.0, 3.1, 8.8, 4.6, 5.8, 4.0, 4.6)
tipoB <- c(16.2, 17.4, 8.5, 15.6, 5.4, 9.8, 14.9, 16.6, 15.9, 5.3, 10.5)
cat("\ Tamanhos amostrais: m =", length(tipoA), ", n =", length(tipoB))
```

```
Tamanhos amostrais: m = 10, n = 11
```

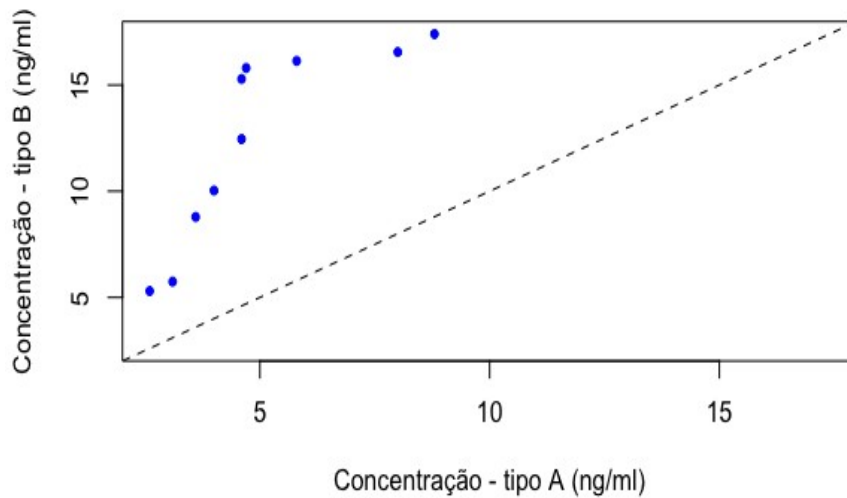
```
# Distribuições empíricas
Fm <- ecdf(tipoA)
Fn <- ecdf(tipoB)
```

<pre># Tipo A tabFm <- cbind(knots(Fm), Fm(knots(Fm))) colnames(tabFm) <- c("x", "Fm(x)") print(tabFm)</pre>	<pre># Tipo B tabFn <- cbind(knots(Fn), Fn(knots(Fn))) colnames(tabFn) <- c("x", "Fn(x)") print(tabFn, digits = 3)</pre>
<pre> x Fm(x) [1,] 2.6 0.1 [2,] 3.1 0.2 [3,] 3.6 0.3 [4,] 4.0 0.4 [5,] 4.6 0.6 [6,] 4.7 0.7 [7,] 5.8 0.8 [8,] 8.0 0.9 [9,] 8.8 1.0</pre>	<pre> x Fn(x) [1,] 5.3 0.0909 [2,] 5.4 0.1818 [3,] 8.5 0.2727 [4,] 9.8 0.3636 [5,] 10.5 0.4545 [6,] 14.9 0.5455 [7,] 15.6 0.6364 [8,] 15.9 0.7273 [9,] 16.2 0.8182 [10,] 16.6 0.9091 [11,] 17.4 1.0000</pre>

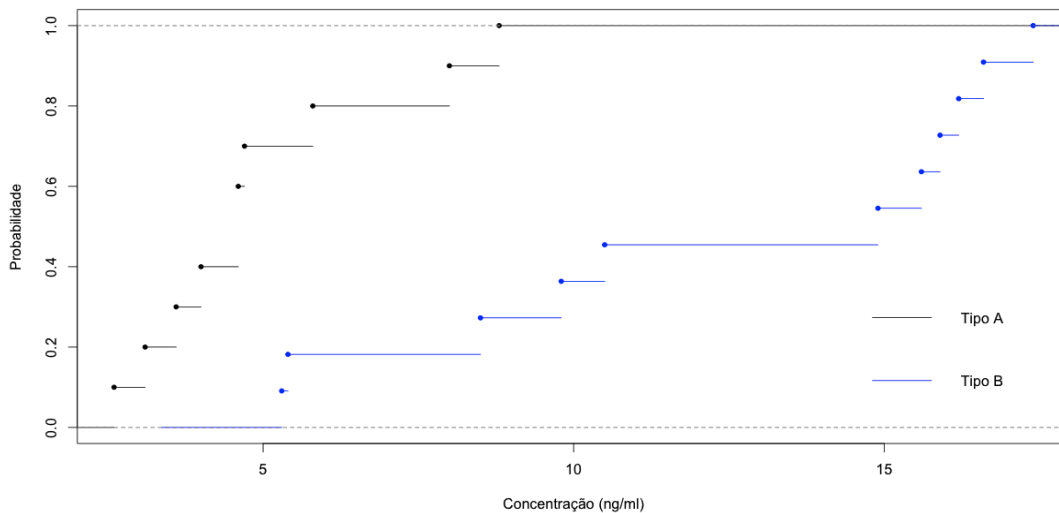
Com o comando `table(tipoA)` notamos que na amostra do tipo A o valor 4,6 foi observado duas vezes. Na função F_m a mudança em $x = 4,6$ é igual a $2 / m = 2 / 10 = 0,2$.

Como as amostras não são de dados pareados, não é possível obter o gráfico de dispersão. Entretanto, com a função `qqplot` é possível obter o gráfico de dispersão dos quantis. Neste exemplo, $m < n$ ($10 < 11$). Internamente, a função `qqplot` aproxima 10 (= m) quantis da variável `tipoB` por interpolação linear (função `approx`).

```
## Gráfico de quantis (QQ)
fAB <- range(tipoA, tipoB)
qqplot(tipoA, tipoB, pch = 20, xlim = fAB, ylim = fAB,
       xlab = "Concentração - tipo A (ng/ml)",
       ylab = "Concentração - tipo B (ng/ml)", col = "blue")
abline(0, 1, lty = 2)
```



```
## Gráficos das distribuições empíricas
plot(Fm, main = "", pch = 20, xlim = range(tipoA, tipoB),
     xlab = "Concentração (ng/ml)", ylab = "Probabilidade")
lines(Fn, col = "blue", pch = 20)
legend("bottomright", c("Tipo A", "Tipo B"), lty = 1,
      col = c("black", "blue"), bty = "n")
```



```
## Cálculo das diferenças
conc <- sort(unique(c(tipoA, tipoB)))
Fmconc <- Fm(conc)
Fnconc <- Fn(conc)
absdif <- abs(Fmconc - Fnconc)
calculo <- cbind(conc, Fmconc, Fnconc, absdif)
colnames(calculo) <- c("Concentração", "Fm", "Fn", "|Fm - Fn|")
print(calculo, digits = 3)
```

	Concentração	Fm	Fn	Fm - Fn
[1,]	2.6	0.1	0.0000	0.1000
[2,]	3.1	0.2	0.0000	0.2000
[3,]	3.6	0.3	0.0000	0.3000
[4,]	4.0	0.4	0.0000	0.4000
[5,]	4.6	0.6	0.0000	0.6000
[6,]	4.7	0.7	0.0000	0.7000
[7,]	5.3	0.7	0.0909	0.6091
[8,]	5.4	0.7	0.1818	0.5182
[9,]	5.8	0.8	0.1818	0.6182
[10,]	8.0	0.9	0.1818	0.7182
[11,]	8.5	0.9	0.2727	0.6273
[12,]	8.8	1.0	0.2727	0.7273
[13,]	9.8	1.0	0.3636	0.6364
[14,]	10.5	1.0	0.4545	0.5455
[15,]	14.9	1.0	0.5455	0.4545
[16,]	15.6	1.0	0.6364	0.3636
[17,]	15.9	1.0	0.7273	0.2727
[18,]	16.2	1.0	0.8182	0.1818
[19,]	16.6	1.0	0.9091	0.0909
[20,]	17.4	1.0	1.0000	0.0000

Os valores únicos (conc) poderiam ser obtidos com o comando

```
conc <- sort(c(knots(Fm), knots(Fn)))
```

```
# Cálculo de Dmn
```

```
Dmn <- max(calculo[, "|Fm - Fn|"])
```

```
cat("\n Dmn = ", Dmn)
```

```
Dmn = 0.7272727
```

```
## Gráficos das distribuições empíricas e Dmn
```

```
(xmax <- conc[which.max(calculo[, "|Fm - Fn|"])])
```

```
[1] 8.8
```

```
plot(Fm, main = "", pch = 20, xlim = range(tipoA, tipoB), xlab = "Concentração (ng/ml)", ylab = "Probabilidade")
```

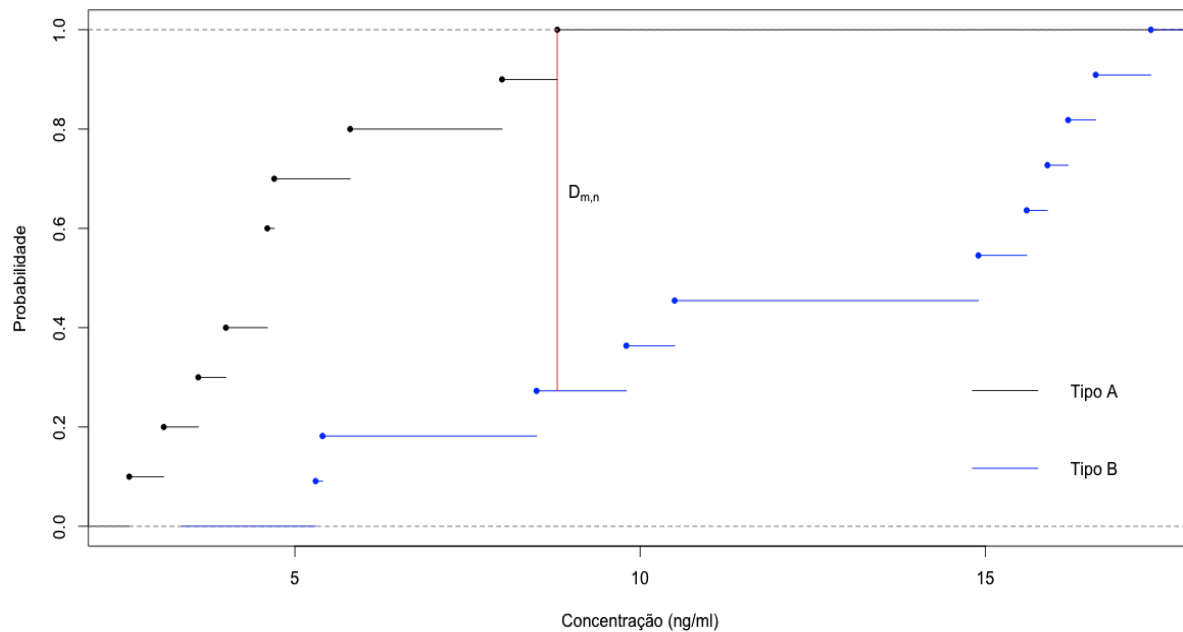
```
lines(Fn, col = "blue", pch = 20)
```

```
legend("bottomright", c("Tipo A", "Tipo B"), lty = 1,
```

```
col = c("black", "blue"), bty = "n")
```

```
segments(xmax, Fm(xmax), xmax, Fn(xmax), col = "red")
```

```
text(locator(1), expression(D[m][","][n]))
```



```
## Teste KS
ks.test(tipoA, tipoB)
```

Two-sample Kolmogorov-Smirnov test

```
data: tipoA and tipoB
D = 0.7273, p-value = 0.007844
alternative hypothesis: two-sided
```

```
Warning message:
In ks.test(tipoA, tipoB) : cannot compute exact p-values with ties
```

Com um nível de significância de 5%, de acordo com o teste de Kolmogorov-Smirnov concluímos que há diferença significativa entre as concentrações para indivíduos dos tipos A e B ($D_{m,n} = 0.7273$; $p = 0,0078$).

Nota 1. Teste a hipótese de que a concentração em indivíduos do tipo A é estocasticamente menor do que a concentração em indivíduos do tipo B.