Teste de bondade de ajuste

Hipótese nula simples

2023

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

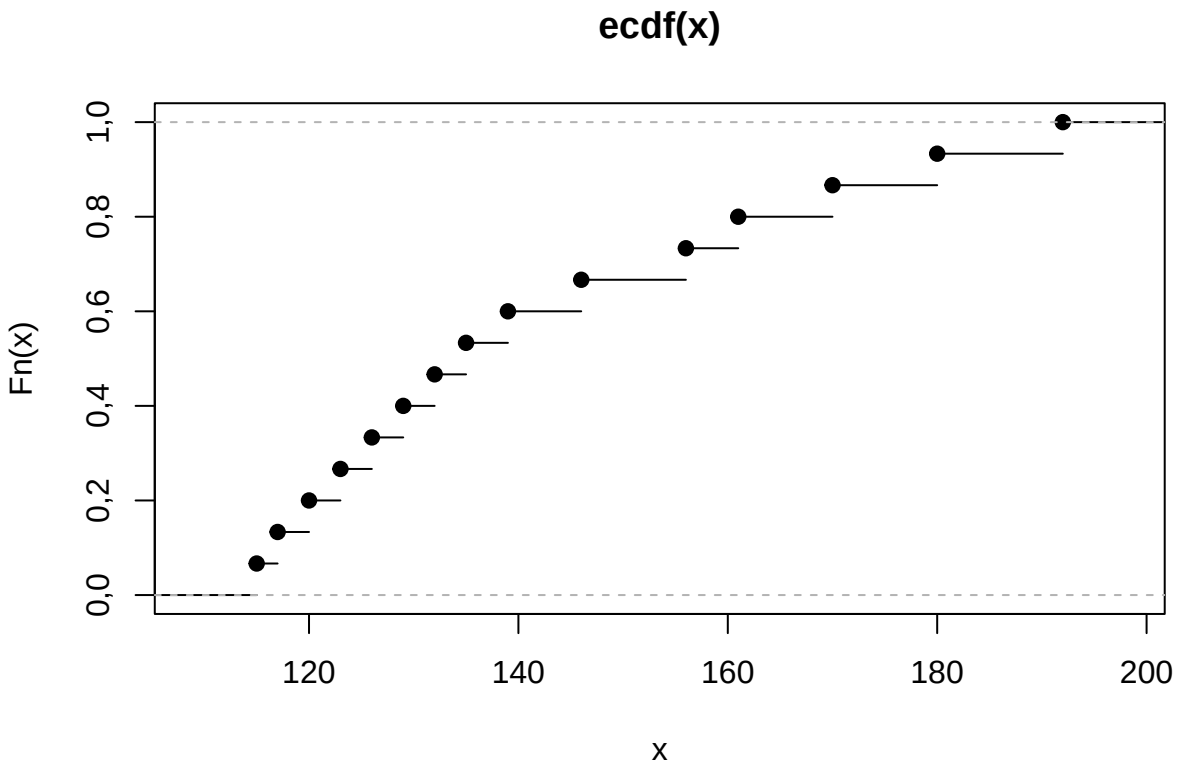
As funções `ks.test` (pacote `stats`), `cvm.test` e `ad.test` (pacote `goftest`) da linguagem R são usadas.

Função distribuição empírica

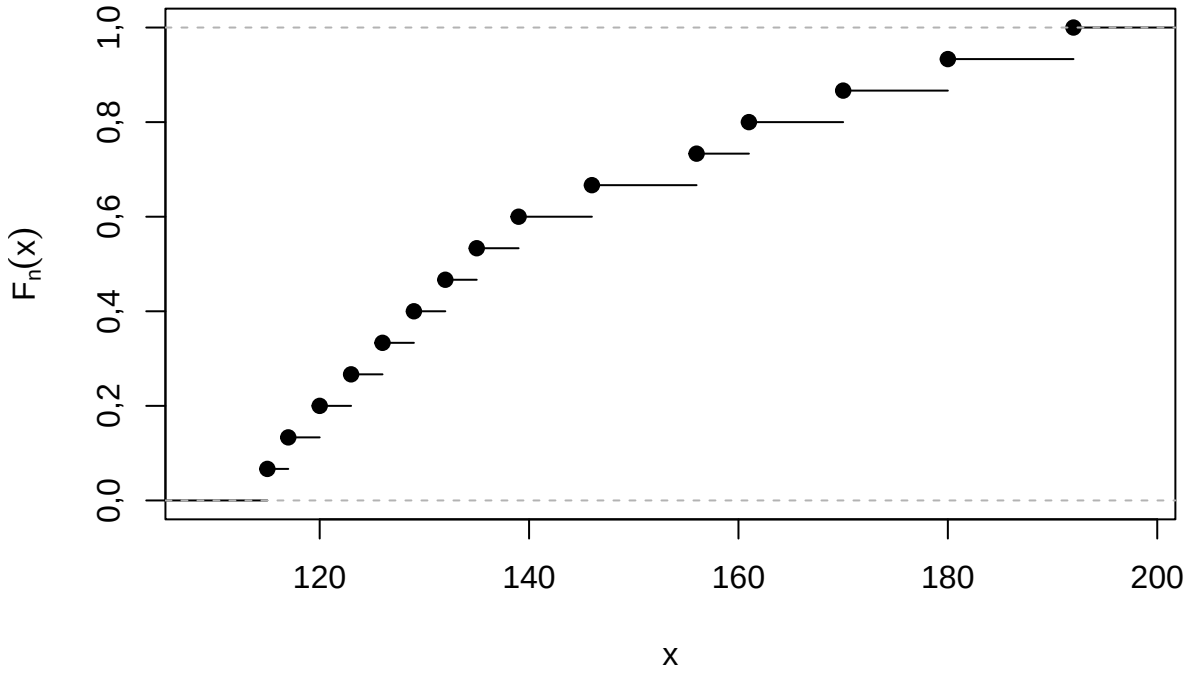
```
# Dados
x <- c(126, 120, 117, 132, 146, 192, 180, 161, 156, 135, 129, 115, 170,
       139, 123)
cat("n = ", length(x))
```

```
## n =  15
```

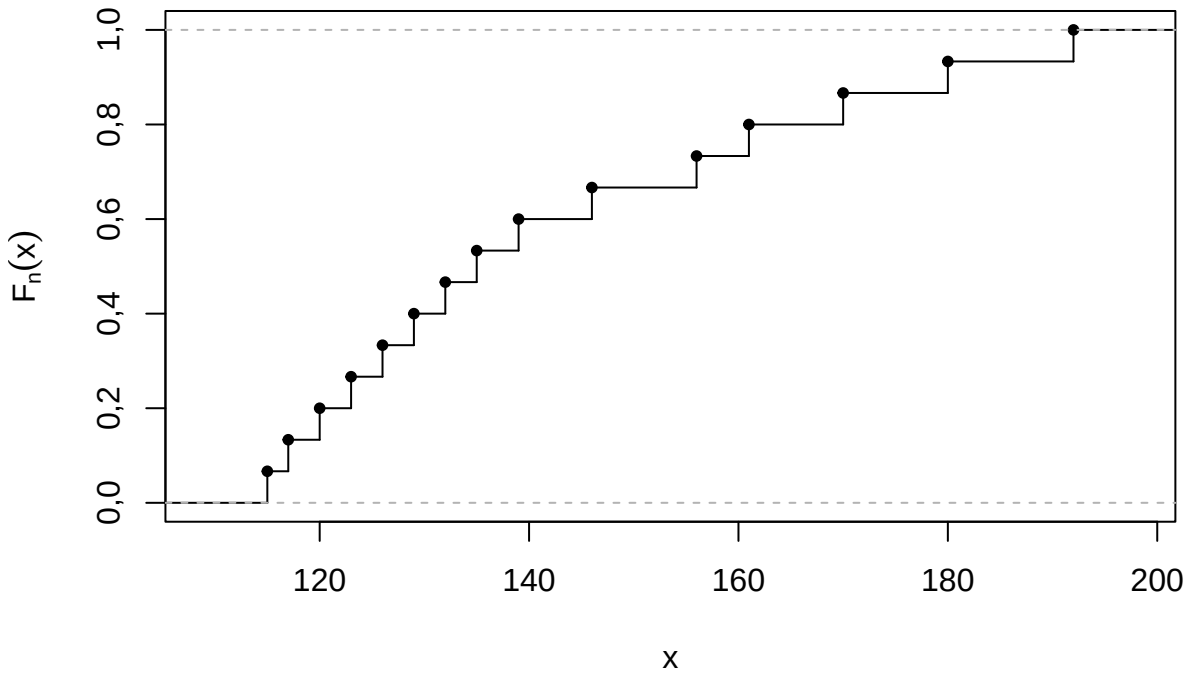
```
# Forma mais simples
plot(ecdf(x))
```



```r
# Mudando alguns argumentos
plot(ecdf(x) , main = "", ylab = expression(F[n](x)))
```

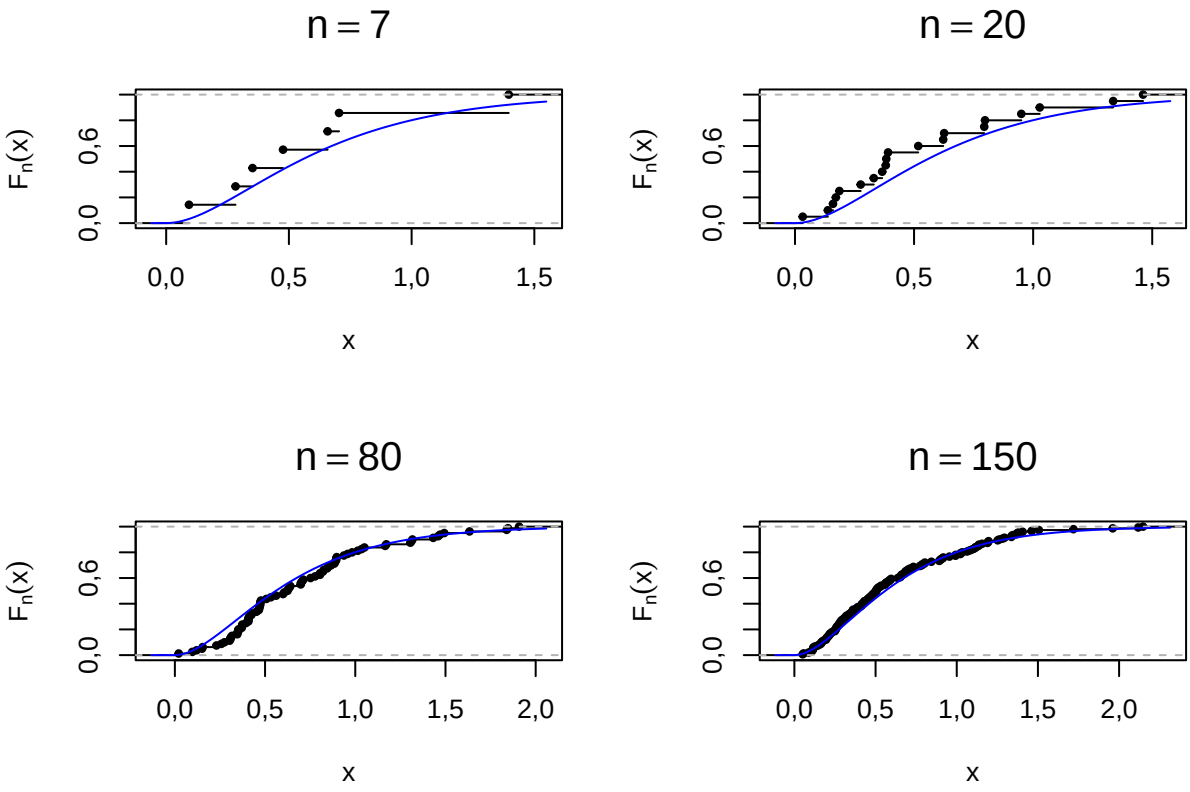



```r
plot(ecdf(x) , main = "", ylab = expression(F[n](x)), pch = 20,
     verticals = TRUE)
```




```r
# Exemplo com diferentes tamanhos de amostra
n <- c(7, 20, 80, 150)
par(mfrow = c(2, 2))
for (tamanho in n) {
   dados <- rgamma(tamanho, shape = 2, rate = 3)
```

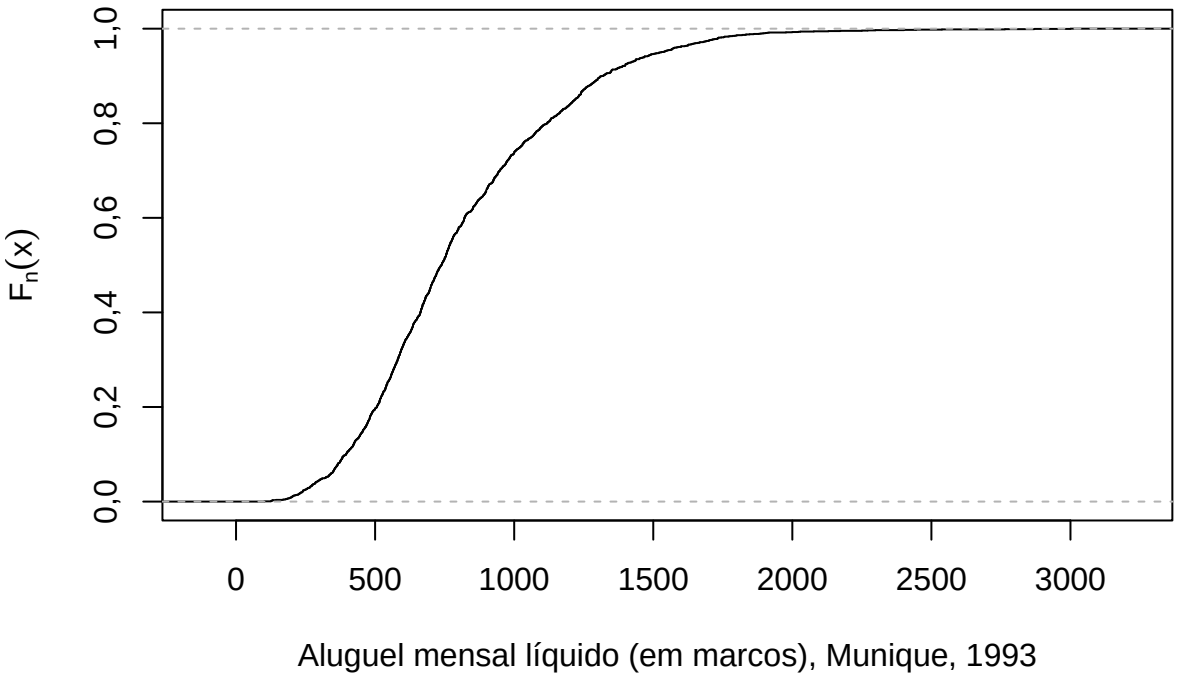
```
   plot(ecdf(dados) , main = bquote(n == .(tamanho)),
        ylab = expression(F[n](x)), pch = 20, cex.main = 1.5)
   curve(pgamma(x, shape = 2, rate = 3), add = TRUE, col = "blue")
}
```




```
# Exemplo com dados de aluguel em Munique
library(gamlss.data)
data("rent")
cat("n =", length(rent$R))
```

```
## n = 1969
```

```
plot(ecdf(rent$R), main = "", pch = 20, ylab = expression(F[n](x)),
xlab = "Aluguel mensal líquido (em marcos), Munique, 1993")
```



Testes

As funções `cvm.test` (teste de Cramér-von Mises) e `ad.test` (teste de Anderson-Darling) estão implementadas no pacote `goftest`.

```
library("goftest")
```

```
# "Dados"
# X ~ gama(forma = f0, taxa = t0)
set.seed(14934)
f0 <- 2
t0 <- 1.5
n <- 45
dados <- rgamma(n, shape = f0, rate = t0)
```

Nas funções `ks.test`, `cvm.test` e `ad.test`, a hipótese nula deve ser especificada em termos de uma função distribuição acumulada (`pnorm`, `pweibull`, ...) e seus respectivos parâmetros.

```
# H0: X ~ gama(forma = f0, taxa = t0)
# Default: H1 bilateral e valor-p exato
(tks <- ks.test(dados, "pgamma", shape = f0, rate = t0))
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  dados
## D = 0,088627, p-value = 0,8408
## alternative hypothesis: two-sided
```

```
cvm.test(dados, null = "pgamma", shape = f0, rate = t0)
```

```
##
##  Cramer-von Mises test of goodness-of-fit
```

```
##  Null hypothesis: Gamma distribution
##  with parameters shape = 2, rate = 1,5
##  Parameters assumed to be fixed
##
## data:  dados
## omega2 = 0,06165, p-value = 0,8061
```

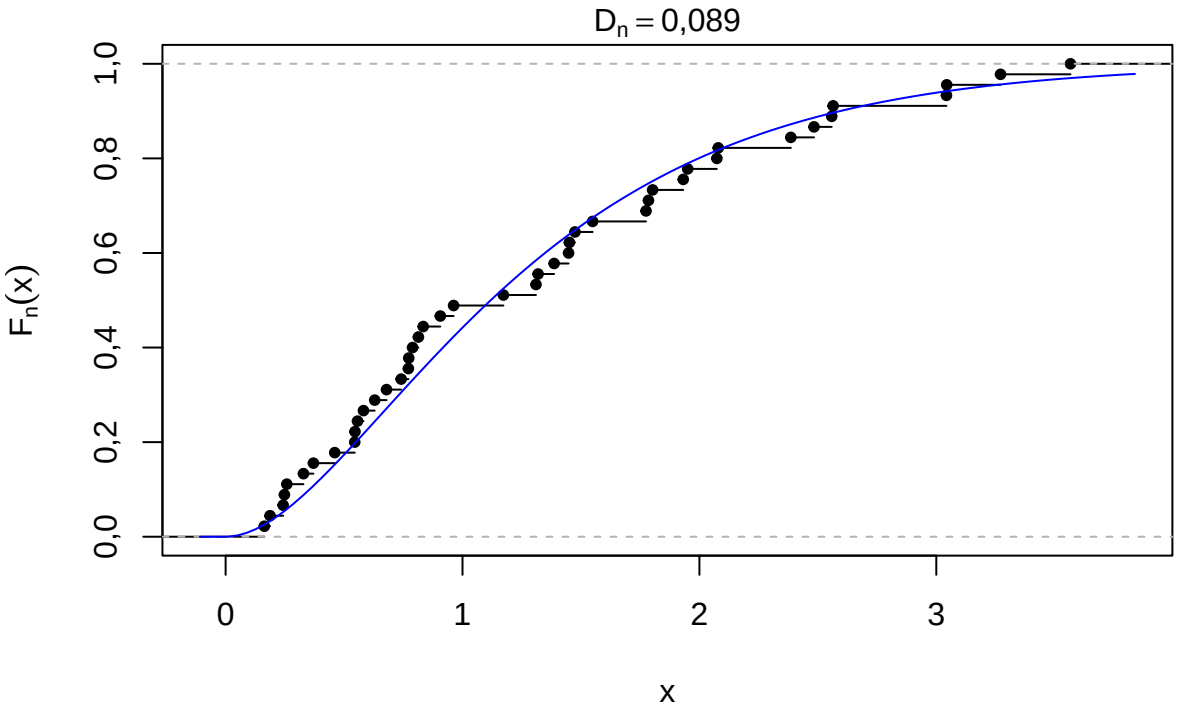
```
ad.test(dados, null = "pgamma", shape = f0, rate = t0)
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Null hypothesis: Gamma distribution
##  with parameters shape = 2, rate = 1,5
##  Parameters assumed to be fixed
##
## data:  dados
## An = 0,37963, p-value = 0,868
```

```
# valor-p aproximado (teste KS)
ks.test(dados, "pgamma", shape = f0, rate = t0, exact = FALSE)
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  dados
## D = 0,088627, p-value = 0,8714
## alternative hypothesis: two-sided
```

```
# Gráficos
plot(ecdf(dados) , main = "", ylab = expression(F[n](x)), pch = 20)
curve(pgamma(x, shape = f0, rate = t0), add = TRUE, col = "blue")
mtext(bquote(D[n] == .(round(tks$statistic, digits = 3))))
```



```r
# H0: X ~ normal(média = f0 / t0, variância = f0 / t0^2)
# Normal com mesma média e mesma variância da dist. gama
# Default: H1 bilateral e valor-p exato
(tksn <- ks.test(dados, "pnorm", mean = f0 / t0, sd = sqrt(f0 / t0^2)))
```

```
##
##  Exact one-sample Kolmogorov-Smirnov test
##
## data:  dados
## D = 0,14619, p-value = 0,2645
## alternative hypothesis: two-sided
```

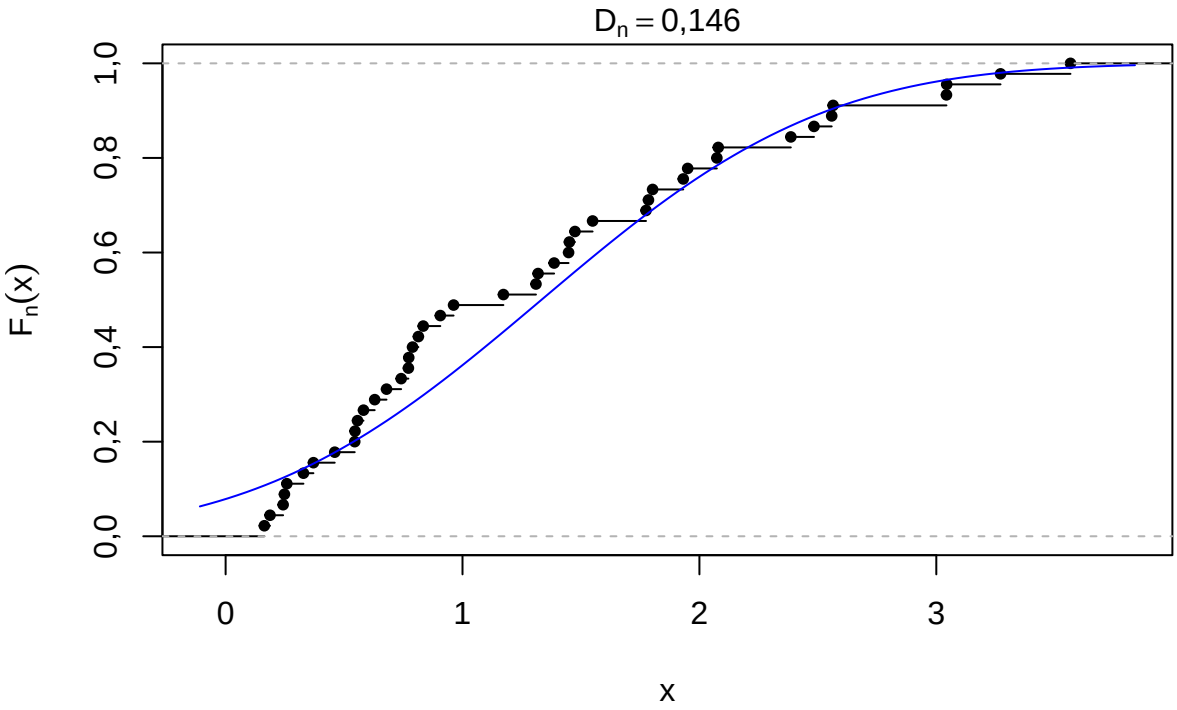
```r
cvm.test(dados, "pnorm", mean = f0 / t0, sd = sqrt(f0 / t0^2))
```

```
##
##  Cramer-von Mises test of goodness-of-fit
##  Null hypothesis: Normal distribution
##  with parameters mean = 1,33333333333333, sd =
##  0,942809041582063
##  Parameters assumed to be fixed
##
## data:  dados
## omega2 = 0,16169, p-value = 0,3568
```

```r
ad.test(dados, "pnorm", mean = f0 / t0, sd = sqrt(f0 / t0^2))
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Null hypothesis: Normal distribution
##  with parameters mean = 1,33333333333333, sd =
##  0,942809041582063
##  Parameters assumed to be fixed
##
## data:  dados
## An = 1,0192, p-value = 0,3469
```

```r
plot(ecdf(dados) , main = "", ylab = expression(F[n](x)), pch = 20)
curve(pnorm(x, mean = f0 / t0, sd = sqrt(f0 / t0^2)), add = TRUE,
  col = "blue")
mtext(bquote(D[n] == .(round(tksn$statistic, digits = 3))))
```



Nota 1. Refaça o teste da hipótese de normalidade aumentando o valor do parâmetro de forma (`f0`). Surpresa?