Testes de normalidade

2023

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

O pacote `MASS` em R contém a função `fitdistr`, que será usada para obter as estimativas dos parâmetros da distribuição normal (e seus respectivos erros padrão). Testes de normalidade estão implementados no pacote `nortest`.

```r
library(MASS)
library(nortest)
```

Neste exemplo a hipótese nula é composta.

```r
## Dados
# Densidade de amostras de concreto, em kg/m^3
# Kottegoda & Rosso (1997, p. 292)

dados <- c(2429, 2469, 2448, 2436, 2454, 2411, 2441, 2445, 2444, 2447,
           2448, 2437, 2488, 2445, 2435, 2428, 2455, 2471, 2456, 2441,
           2456, 2425, 2446, 2415, 2472, 2435, 2447, 2458, 2427, 2436,
           2450, 2449, 2457, 2437, 2433, 2436, 2427, 2454, 2473, 2436)

cat("\n Número de observações (n):", length(dados))
```

```
##
##  Número de observações (n): 40
```

```r
## Estimativas dos parâmetros
(estpar <- fitdistr(dados, "normal"))
```

```
##       mean           sd
##   2444,925000     15,793017
##  (   2,497095) (   1,765713)
```

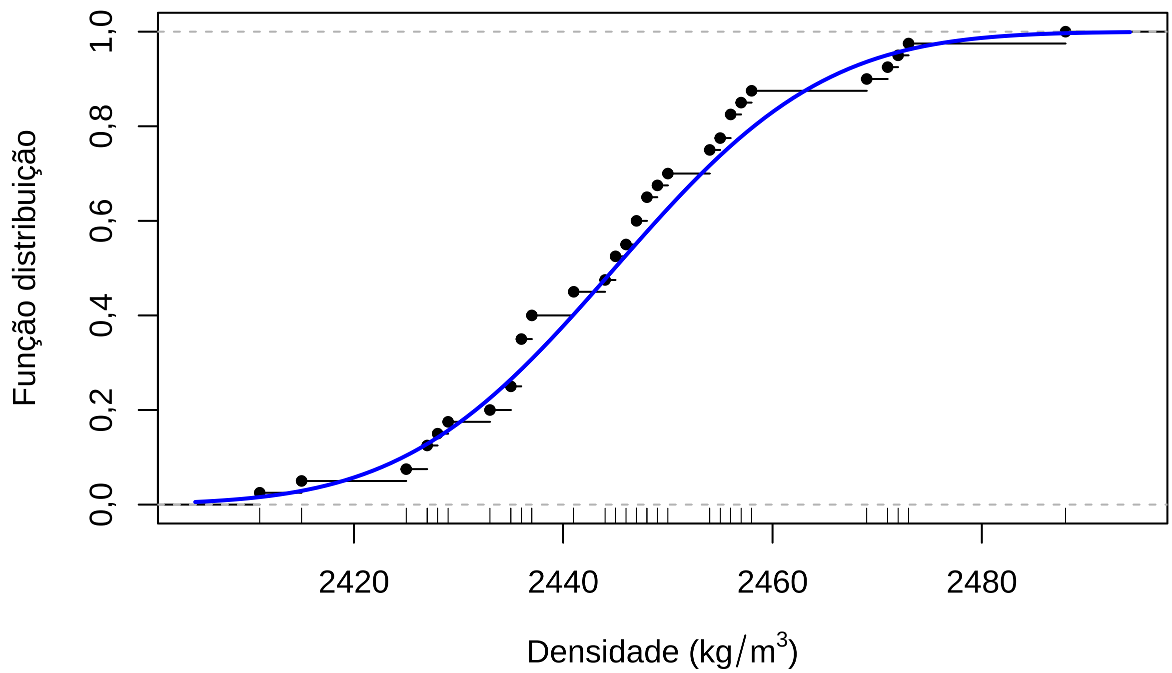
No resultado acima, `mean` e `sd` são as estimativas dos parâmetros da distribuição normal, que são a média e o desvio padrão, sendo que os valores entre parênteses são os erros padrão.

```r
names(estpar)
```

```
## [1] "estimate" "sd"       "vcov"     "n"        "loglik"
```

Nota 1. Procure entender o significado de cada um dos componentes acima.

```r
## Função distribuição empírica e distribuição normal
plot(ecdf(dados), main = "", pch = 20,
    xlab = expression(paste("Densidade (", kg/m^3, ")")),
    ylab = "Função distribuição")
rug(dados)
curve(pnorm(x, mean = estpar$estimate[1], sd = estpar$estimate[2]),
      col = "blue", add = TRUE, lwd = 2)
```




```r
## Teste de Anderson-Darling
ad.test(dados)
```

```
##
##  Anderson-Darling normality test
##
## data:  dados
## A = 0,3877, p-value = 0,3714
```

```r
## Teste de Cramér-von Mises
cvm.test(dados)
```

```
##
##  Cramer-von Mises normality test
##
## data:  dados
## W = 0,058591, p-value = 0,3862
```

```r
## Teste de Lilliefors
lillie.test(dados)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  dados
## D = 0,089874, p-value = 0,5734
```

O teste qui-quadrado de Pearson, por *default*, utiliza $k = \lceil 2n^{2/5} \rceil$ classes equiprováveis, sendo que $\lceil b \rceil$ denota o menor inteiro maior do que ou igual a b (função `ceiling` em R). Neste exemplo, $n = 40$ e $k = \lceil 2 \times 40^{2/5} \rceil = 9$, que leva a $9 - 2 - 1 = 6$ graus de liberdade.

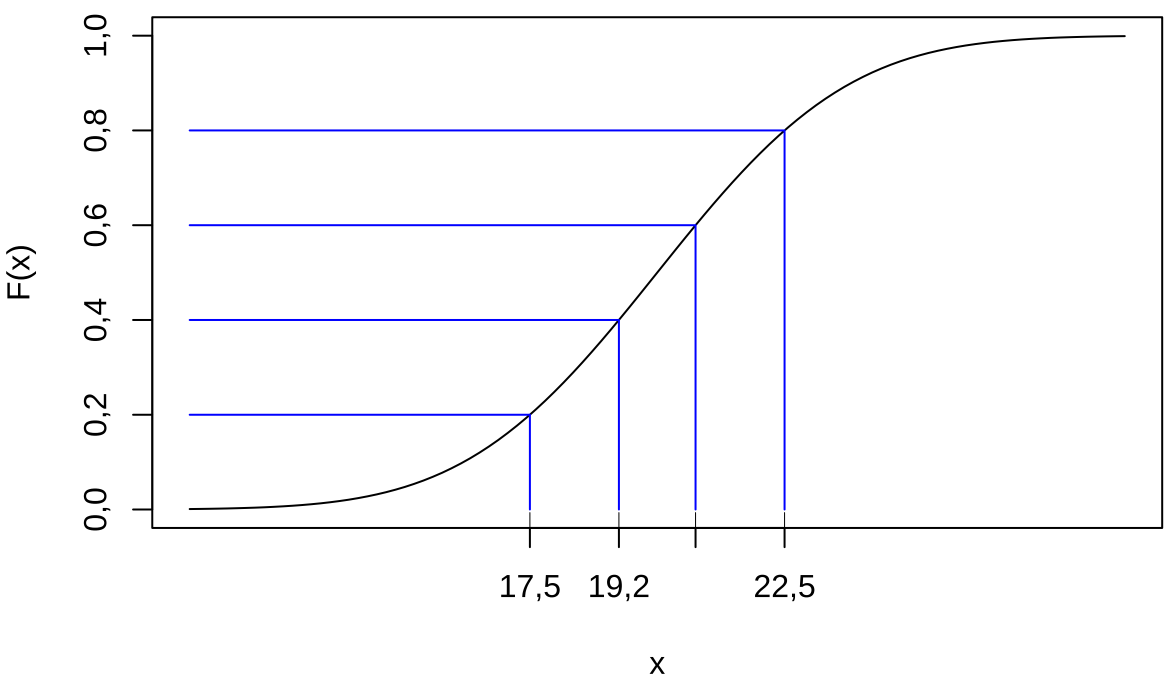
```r
## Teste qui-quadrado de Pearson
pearson.test(dados)
```

```
##
##  Pearson chi-square normality test
```

```
##
## data:  dados
## P = 10,4, p-value = 0,1088
```

O gráfico abaixo mostra os cinco intervalos (probabilidade = 1/5) para uma distribuição normal com média 20 e desvio padrão 3.

```r
m0 <- 20
sd0 <- 3
f0 <- qnorm(0.001, m0, sd0)
f1 <- qnorm(0.999, m0, sd0)
k <- 5
p0 <- (1:(k - 1)) / k
q0 <- qnorm(p0, mean = m0, sd = sd0)
curve(pnorm(x, mean = m0, sd = sd0), from = f0, to = f1, ylab = "F(x)",
      axes = FALSE)
axis(1, at = q0, labels = round(q0, 1))
axis(2)
rug(q0)
segments(q0, 0, q0, p0, col = "blue")
segments(f0, p0, q0, p0, col = "blue")
box()
```



A estatística do teste de Shapiro-Francia é dada pelo quadrado do coeficiente de correlação entre as estatísticas de ordem da amostra e uma aproximação das estatísticas de ordem das distribuição normal padrão.

```r
## Teste de Shapiro-Francia
sf.test(dados)
```

```
##
##  Shapiro-Francia normality test
##
## data:  dados
## W = 0,97249, p-value = 0,3636
```

De acordo com os cinco testes realizados e com um nível de significância de 5%, não se rejeita a hipótese de que os dados de densidade de concreto seguem uma distribuição normal.

Nota 2. O teste de normalidade de Shapiro-Wilk está implementado na função `shapiro.test` do pacote `stats`.

Nota 3. Refaça os testes com as funções do pacote `goftest`.