Teste do sinal

Linguagem R

2023

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

Para calcular o valor-p são utilizadas as funções `binom.test` (teste exato) e `prop.test` (teste aproximado com ou sem correção de continuidade).

Exemplo 1

```
# Tabela 3.11, p. 83, em Hollander & Wolf (1999, 2nd ed.)
# H1 unilateral à direita (teta > teta0)
teta0 <- 175
x <- c(254, 171, 345, 134, 190, 447, 106, 173, 449, 198)
n <- length(x)
cat("\n Tamanho da amostra:", n, "\n")
```

```
##
##  Tamanho da amostra: 10
```

```
Bobs <- sum(x > teta0)
cat("\n Bobs =", Bobs,"\n")
```

```
##
##  Bobs = 6
```

```
# Dist. exata (default: p = 1/2)
(binom.test(Bobs, n, alternative = "greater"))
```

```
##
##  Exact binomial test
##
## data:  Bobs and n
## number of successes = 6, number of trials = 10, p-value =
## 0,377
## alternative hypothesis: true probability of success is greater than 0,5
## 95 percent confidence interval:
##  0,3035372 1,0000000
## sample estimates:
## probability of success
##                    0,6
```

```
# Valor-p com a dist. de B (pbinom)
cat("\n Bobs =", Bobs, "(p =", pbinom(Bobs - 1, n, prob = 0.5, lower.tail = FALSE), ")")
```

```
##
```

```
##  Bobs = 6 (p = 0,3769531 )
```

Pelo resultado acima, concluímos que a mediana não é significativamente maior do que 175 a um nível de significância de 5%.

```r
# Dist. aproximada com correção de continuidade
(prop.test(Bobs, n, p = 0.5, alternative = "greater", correct = TRUE))
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  Bobs out of n, null probability 0.5
## X-squared = 0,1, df = 1, p-value = 0,3759
## alternative hypothesis: true p is greater than 0,5
## 95 percent confidence interval:
##  0,3095345 1,0000000
## sample estimates:
##   p
## 0,6
```

```r
# Dist. aproximada sem correção de continuidade
(prop.test(Bobs, n, p = 0.5, alternative = "greater", correct = FALSE))
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  Bobs out of n, null probability 0.5
## X-squared = 0,4, df = 1, p-value = 0,2635
## alternative hypothesis: true p is greater than 0,5
## 95 percent confidence interval:
##  0,3516386 1,0000000
## sample estimates:
##   p
## 0,6
```

Exemplo 2

```r
# Tabela 3.9, p. 82, em Hollander & Wolf (1999, 2nd ed.)
# H1 bilateral
teta0 <- 18
x <- c(17.4, 17.9, 17.6, 18.1, 17.6, 18.9, 16.9, 17.5, 17.8, 17.4, 24.6, 26.0)
n <- length(x)
cat("\n Tamanho da amostra:", n, "\n")
```

```
##
##  Tamanho da amostra: 12
```

```r
Bobs <- sum(x > teta0)
cat("\n Bobs =", Bobs,"\n")
```

```
##
##  Bobs = 4
```

```r
# Dist. exata
(binom.test(Bobs, n, alternative = "two.side"))
```

```
##
##  Exact binomial test
##
## data:  Bobs and n
## number of successes = 4, number of trials = 12, p-value =
## 0,3877
## alternative hypothesis: true probability of success is not equal to 0,5
## 95 percent confidence interval:
##  0,09924609 0,65112449
## sample estimates:
## probability of success
##              0,3333333
```

```r
# Valor-p com a dist. de B (pbinom)
if (Bobs <= n / 2) {
  vp <- 2 * pbinom(Bobs, n, prob = 0.5)
} else {
  vp <- 2 * pbinom(Bobs - 1, n, prob = 0.5, lower.tail = FALSE,
                   lower.tail = FALSE)
}
cat("\n Bobs =", Bobs, "(p =", vp, ")")
```

```
##
##  Bobs = 4 (p = 0,3876953 )
```

Pelo resultado acima, concluímos que a mediana não é significativamente diferente de 18 a um nível de significância de 5%.

```r
# Estimativa pontual de teta
cat("\n Mediana amostral =", median(x), "\n")
```

```
##
##  Mediana amostral = 17,7
```

Em seguida determinamos as posições dos limites do intervalo de confiança para a mediana. A posição do limite superior é dada pelo menor valor de b tal que

$$\frac{1}{2^n}\sum_{j=b}^{n}\binom{n}{j} \leq \frac{\alpha}{2}.$$

Calculamos $C_\alpha = n + 1 - b$. Em termos da distribuição `binomial`$(n, 1/2)$, a região crítica corresponde à união dos conjuntos $\{0, 1, \ldots, C_\alpha - 1\} = \{0, 1, \ldots, n - b\}$ e $\{b, b + 1, \ldots, n\}$, cada um deles com $n + 1 - b$ elementos (distribuição simétrica em torno de $n/2$). Sendo assim, o intervalo de confiança tem probabilidade de cobertura dada por

$$\frac{1}{2^n}\sum_{j=C_\alpha}^{b-1}\binom{n}{j}.$$

Os valores C_α e b garantem que o teste do sinal para $\mathrm{H}_0 : \theta = \theta_0$ contra $\mathrm{H}_1 : \theta \neq \theta_0$ tem nível de significância $\leq \alpha$. O intervalo de confiança baseado no teste do sinal tem coeficiente de confiança $\geq 1 - \alpha$.

Nota 1. Se q_γ é o quantil correspondente à probabilidade γ, temos que

$$\sum_{j=q_\gamma+1}^{n} P(X = j) \leq \frac{\alpha}{2}.$$

Justifique este resultado.

```r
# IC de 95% para teta
alfa <- 0.05
(b <- qbinom(1 - alfa / 2, n, prob = 0.5) + 1)
```

```
## [1] 10
```

```r
(calfa <- n + 1 - b)
```

```
## [1] 3
```

```r
conf <- sum(dbinom(calfa:(b - 1), n, 0.5))
xs <- sort(x)
cat("\n IC de", round(100 * conf, 1), "% para teta: (", xs[calfa], ",",
    xs[b], ") \n")
```

```
##
##  IC de 96,1 % para teta: ( 17,4 , 18,9 )
```

```r
# Dist. aproximada com correção de continuidade
(prop.test(Bobs, n, p = 0.5, alternative = "two.side", correct = TRUE))
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  Bobs out of n, null probability 0.5
## X-squared = 0,75, df = 1, p-value = 0,3865
## alternative hypothesis: true p is not equal to 0,5
## 95 percent confidence interval:
##  0,1127286 0,6456310
## sample estimates:
##         p
## 0,3333333
```

```r
# Dist. aproximada sem correção de continuidade
(prop.test(Bobs, n, p = 0.5, alternative = "two.side", correct = FALSE))
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  Bobs out of n, null probability 0.5
## X-squared = 1,3333, df = 1, p-value = 0,2482
## alternative hypothesis: true p is not equal to 0,5
## 95 percent confidence interval:
##  0,1381201 0,6093779
## sample estimates:
##         p
## 0,3333333
```

Nota 2. Utilizando os dados do exemplo 2, efetue o teste com $\theta_0 = 18,1$.

Nota 3. Refaça os exemplos usando a função `SIGN.test` do pacote `BSDA`.

Nota 4. Teste as hipóteses dos exemplos 1 e 2 utilizando os testes de Wilcoxon e t de Student. Alguma diferença nas suas conclusões?