

Função chisq.test

São apresentados exemplos do teste qui-quadrado de Pearson de bondade do ajuste. O teste está implementado na função `chisq.test` do pacote `stats` em R. Em todos os exemplos a hipótese nula é simples.

Exemplo 1. Testar se uma distribuição multinomial com probabilidades $(1/3, 1/3, 1/3)$ ajusta bem as frequências $(20, 15, 25)$.

Neste exemplo, $k = 3$ e $n = 20 + 15 + 25 = 60$. Como as probabilidades sob a hipótese nula são iguais a $1/k$, não é necessário informá-las.

```
dados <- c(20, 15, 25)
(ex1 <- chisq.test(dados))
```

Chi-squared test for given probabilities

```
data: dados
X-squared = 2.5, df = 2, p-value = 0.2865.
```

O valor da estatística de teste é $Q = 2,5$ e está armazenado em `ex1$statistic`. Com dois graus de liberdade ($k - 1 = 2$), obtemos $\text{valor-}p = 0,2865$, que pode ser calculado como

```
pchisq(ex1$statistic, df = length(dados) - 1, lower.tail = FALSE)
```

ou

```
pchisq(ex1$statistic, df = ex1$parameter, lower.tail = FALSE)
```

```
0.2865048
```

Exemplo 2. Testar se uma distribuição multinomial com probabilidades $(1/4, 1/8, 5/8)$ ajusta bem as frequências $(14, 25, 81)$.

Neste exemplo, $k = 3$ e $n = 120$. Como as probabilidades sob a hipótese nula são diferentes de $1/k$, devemos informar estas probabilidades com o argumento `p`.

```
dados <- c(14, 25, 81)
prob0 <- c(1/4, 1/8, 5/8)
(ex12 <- chisq.test(dados, p = prob0))
```

Chi-squared test for given probabilities

```
data: dados
X-squared = 15.68, df = 2, p-value = 0.0003937
```

Exemplo 3. Testar se uma distribuição multinomial com probabilidades proporcionais a (2, 5, 3, 8) ajusta bem as frequências (19, 62, 31, 105).

Neste exemplo, $k = 4$ e $n = 217$. As probabilidades em si não foram fornecidas, mas são iguais às constantes de proporcionalidade divididas pela sua soma. Podemos informar diretamente as constantes de proporcionalidade, bastando especificar o argumento `rescale.p` como `TRUE`.

```
dados <- c(19, 62, 31, 105)
prop0 <- c(2, 5, 3, 8)
chisq.test(dados, p = prop0, rescale.p = TRUE)
```

Chi-squared test for given probabilities

```
data: dados
X-squared = 2.6297, df = 3, p-value = 0.4523
```

O valor- p do teste pode ser aproximado por simulações de Monte Carlo (especificamos `simulate.p.value = TRUE`). No comando a seguir a aproximação utiliza $B = 5000$ réplicas do vetor de frequências obtidas sob H_0 .

```
(chisq.test(dados, p = prop0, rescale.p = TRUE, simulate.p.value =
  TRUE, B = 5000))
```

Chi-squared test for given probabilities with simulated p-value (based on 5000 replicates)

```
data: dados
X-squared = 2.6297, df = NA, p-value = 0.4393
```

Nota 1. Por que não há informação sobre os graus de liberdade ($df = NA$) na linha acima?

Nota 2. Refaça os exemplos 1 e 2 utilizando uma aproximação de Monte Carlo para o valor- p .

Nota 3. Refaça os três exemplos com a estatística $-2 \log(\Lambda)$ no lugar da estatística Q de Pearson.