

# Teste qui-quadrado de Pearson

2023

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

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 linguagem R.

## Hipótese nula 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 valor- $p = 0,2865$ , que pode ser calculado como

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

```
## X-squared  
## 0,2865048
```

ou

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

```
## X-squared  
## 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
```

## Hipótese nula composta

**Exemplo 4.** Deve ser testada a hipótese de que quatro contagens seguem uma distribuição multinomial com probabilidades

$$p_1 = \theta^2, \quad p_2 = \theta(1 - \theta), \quad p_3 = \theta(1 - \theta) \quad \text{e} \quad p_4 = (1 - \theta)^2, \quad (1)$$

em que  $0 < \theta < 1$ . Em uma amostra de  $n = 215$  observações as frequências observadas são  $n_1 = 19$ ,  $n_2 = 62$ ,  $n_3 = 90$  e  $n_4 = 44$ .

A estimativa de máxima verossimilhança de  $\theta$  supondo que a hipótese (1) é verdadeira é dada por  $(2n_1 + n_2 + n_3)/(2n)$ .

**Nota 1.** Prove o resultado acima.

```
## Dados e estimativa de teta
dados <- c(19, 62, 90, 44)
n <- sum(dados)

emvteta = (2 * dados[1] + dados[2] + dados[3]) / (2 * n)
cat("\n dados:", dados)
```

```
##
## dados: 19 62 90 44
cat("\n n =", n, "\n EMV de teta =", emvteta)
```

```
##
## n = 215
## EMV de teta = 0,4418605
```

Em seguida é realizado o teste da hipótese em (1), notando que um parâmetro  $\theta$  é estimado e o número de graus de liberdade é  $k - 1 - 1 = 4 - 1 - 1 = 2$ .

```
## Frequências esperadas estimadas
piteta <- c(emvteta^2, emvteta * (1 - emvteta), emvteta * (1 - emvteta),
           (1 - emvteta)^2)
esp <- n * piteta
cat("\n Frequências esperadas estimadas:", round(esp, digits = 2))
```

```
##
## Frequências esperadas estimadas: 41,98 53,02 53,02 66,98
```

```
## Estatística de teste
Q <- sum((dados - esp)^2 / esp)
cat("\n Q =", Q, "(p =", pchisq(Q, 2, lower.tail = FALSE), ")")
```

```
##
## Q = 47,7652 (p = 4,24539e-11 )
```

O cálculo da estatística  $Q$  de Pearson pode ser realizado com a função `chisq.test` em R utilizando o vetor de probabilidades estimadas supondo (1) verdadeira (`piteta`). Observe que o valor- $p$  não é correto, pois refere-se ao teste da hipótese simples com  $k - 1 = 4 - 1 = 3$  graus de liberdade (`df = 3`).

```
chisq.test(dados, p = piteta)
```

```
##
## Chi-squared test for given probabilities
##
## data: dados
## X-squared = 47,765, df = 3, p-value = 2,389e-10
```

O resultado da função `chisq.test` inclui vários componentes.

```
tp <- chisq.test(dados, p = piteta)
names(tp)
```

```
## [1] "statistic" "parameter" "p.value" "method" "data.name"
## [6] "observed" "expected" "residuals" "stdres"
```

```
cat("\n Frequências esperadas estimadas:", round(tp$expected, digits = 2))
```

```
##
## Frequências esperadas estimadas: 41,98 53,02 53,02 66,98
```

```
cat("\n Q =", tp$statistic)
```

```
##
## Q = 47,7652
```