

Teste qui-quadrado

São apresentados exemplos com a função `chisq.test` do pacote `stats` em R.

A função `chisq.test` pode ser utilizada para testar a bondade do ajuste a uma distribuição multinomial. Também pode ser utilizada para testar a independência de duas variáveis a partir das contagens em uma tabela de contingências bidimensional.

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

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

```
options(OutDec = ",") # separador decimal é ","
```

```
x <- c(20, 15, 25)
(ex11 <- chisq.test(x))
```

```
Chi-squared test for given probabilities
```

```
data: x
X-squared = 2,5, df = 2, p-value = 0,2865
```

O valor da estatística de teste é $X^2 = 2,5$ e está armazenado em `ex11$statistic`. Com dois graus de liberdade ($N - 1 = 2$), obtemos valor- $p = 0,2865$, que pode ser calculado como

```
pchisq(ex11$statistic, df = length(x) - 1, lower.tail = FALSE)
```

ou

```
pchisq(ex11$statistic, df = ex11$parameter, lower.tail = FALSE)
```

```
0,2865048
```

Nos dois comandos acima, `lower.tail = FALSE` significa que foi calculada a probabilidade na cauda direita (*upper*).

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

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

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

Chi-squared test for given probabilities

```
data: x
X-squared = 15,68, df = 2, p-value = 0,0003937
```

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

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

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

Chi-squared test for given probabilities

```
data: x
X-squared = 2,6297, df = 3, p-value = 0,4523
```

Exemplo 4. A partir dos dados (obtidos sem que as margens fossem fixadas)

```
n <- as.table(rbind(c(762, 327, 468), c(484, 239, 477)))
dimnames(n) <- list(gender = c("M", "F"),
                    party = c("Democrat", "Independent", "Republican"))
n
```

```
      party
gender Democrat Independent Republican
M           762           327           468
F           484           239           477
```

testar a hipótese de independência entre as variáveis *gender* e *party*

Os dados compõem uma tabela 2×3 e o tamanho da amostra é $n = \text{sum}(n) = 2757$. Adotando *gender* como variável explicativa, os gráficos de barras da Figura 1, obtidos com os comandos

```
tab14 <- prop.table(n, margin = 1) * 100
library(lattice)
barchart(tab14, xlab = "Percentage", ylab = "Gender", stack = FALSE,
scale = list(cex = 1.5), auto.key = list(space = "top", columns = 3))
```

sugerem que há dependência entre as variáveis (por quê?). Realizando o teste de independência com a estatística X^2 ,

```
(ex14 <- chisq.test(n))
```

Pearson's Chi-squared test

```
data: n
```

```
X-squared = 30,0701, df = 2, p-value = 2,954e-07
```

encontramos evidências contra a hipótese nula de independência. Vale ressaltar que a função `chisq.test` neste exemplo serve para testar uma hipótese diferente daquelas dos exemplos anteriores.

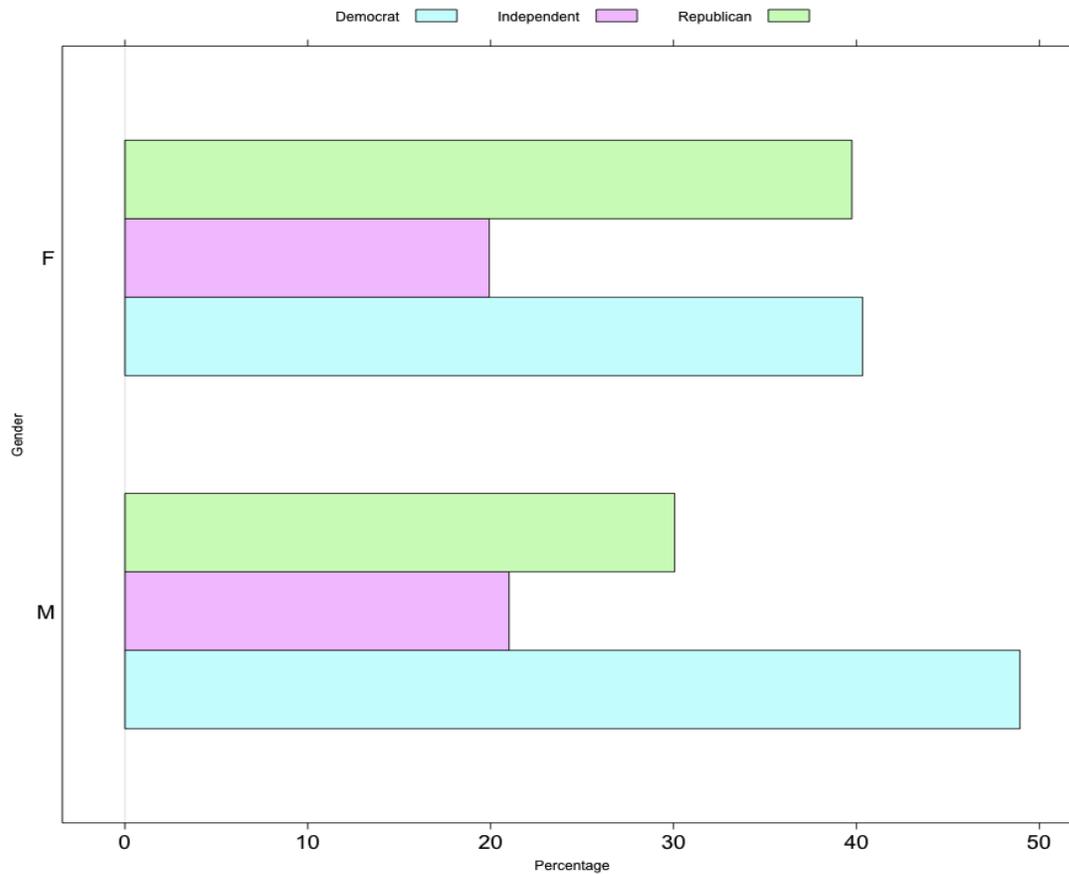


Figura 1. Gráficos de barras do exemplo 4.

As frequências esperadas estimadas sob H_0 são

```
ex14$expected
```

```
party
gender Democrat Independent Republican
M 703.6714 319.6453 533.6834
F 542.3286 246.3547 411.3166
```

A estatística de teste G^2 (baseada na razão de verossimilhanças) tem valor

```
(G2 <- 2 * sum(n * (log(n) - log(ex14$expected))))
```

```
[1] 30,01669
```

bastante próximo ao valor de X^2 .