

Teste de Cochran-Mantel-Haenszel

Apresentamos um exemplo com a função `mantelhaen.test` do pacote `stats` em R.

Os dados do exemplo podem ser encontrados na Tabela 3.3, p. 60, do livro *Introduction to Categorical Data Analysis* (Agresti, A., 1996, Wiley: New York). Os dados envolvem oito estudos sobre a relação entre indicador de ser fumante ($X = 1$: sim; $X = 2$: não) e câncer de pulmão ($Y = 1$: sim; $Y = 2$: não) em cidades chinesas. A variável de controle (Z) é a cidade onde cada estudo foi realizado, com $K = 8$ níveis.

Os dados são organizados em um arranjo (*array*) tridimensional, sendo que a variável de controle (Z : cidade) ocupa a terceira dimensão. O argumento `dim = c(2, 2, 8)` especifica que a tabela é $2 \times 2 \times 8$.

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

dados <- array(c(126, 35, 100, 61,
                908, 497, 688, 807,
                913, 336, 747, 598,
                235, 58, 172, 121,
                402, 121, 308, 215,
                182, 72, 156, 98,
                60, 11, 99, 43,
                104, 21, 89, 36),
              dim = c(2, 2, 8),
              dimnames = list(ifumante = c("Sim", "Não"),
                              icancer = c("Sim", "Não"),
                              cidade = c("Beijing", "Shanghai", "Shenyang", "Nanjing", "Harbin",
                                          "Zhengzhou", "Taiyuan", "Nanchang")))
```

Cada linha na chamada da função `array` corresponde a uma tabela parcial, que é preenchida por colunas. O número total de participantes é 8419, obtido com `sum(dados)`.

`dados`

```
, , cidade = Beijing      , , cidade = Nanjing      , , cidade = Taiyuan
  icancer
ifumante Sim Não        ifumante Sim Não        ifumante Sim Não
  Sim 126 100           Sim 235 172           Sim 60 99
  Não 35 61            Não 58 121           Não 11 43

, , cidade = Shanghai    , , cidade = Harbin      , , cidade = Nanchang
  icancer
ifumante Sim Não        ifumante Sim Não        ifumante Sim Não
  Sim 908 688           Sim 402 308           Sim 104 89
  Não 497 807          Não 121 215           Não 21 36

, , cidade = Shenyang    , , cidade = Zhengzhou
  icancer
ifumante Sim Não        ifumante Sim Não
  Sim 913 747           Sim 182 156
  Não 336 598          Não 72 98
```

Nota 1. Interprete o resultado do comando `apply(dados, 3, sum)`.

Estimativas das razões de chances nas tabelas parciais podem ser obtidas com a função `oddsratio` do pacote `vcd`. A variável de controle Z é especificada com o argumento `stratum` (estrato).

```
library(vcd)
(rcparcial = oddsratio(dados, log = FALSE, stratum = 3))

      Beijing  Shanghai  Shenyang  Nanjing
2,196000  2,142962  2,175265  2,850341

      Harbin  Zhengzhou  Taiyuan  Nanchang
2,319148  1,587963  2,369146  2,003210
```

Para estes dados, cada contagem não é adicionada de 0,5, pois nenhuma das contagens é igual a 0 (por *default*, se pelo menos uma das contagens for nula, `correct = TRUE`). O argumento `log = FALSE` indica intervalos para a razão de chances (por *default*, `log = TRUE`, que significa intervalos para o logaritmo da razão de chances). O argumento `stratum = 3` indica que a variável de controle ocupa a dimensão 3 nos dados tridimensionais.

Intervalos de confiança assintóticos de 95% são obtidos com a função `confint`.

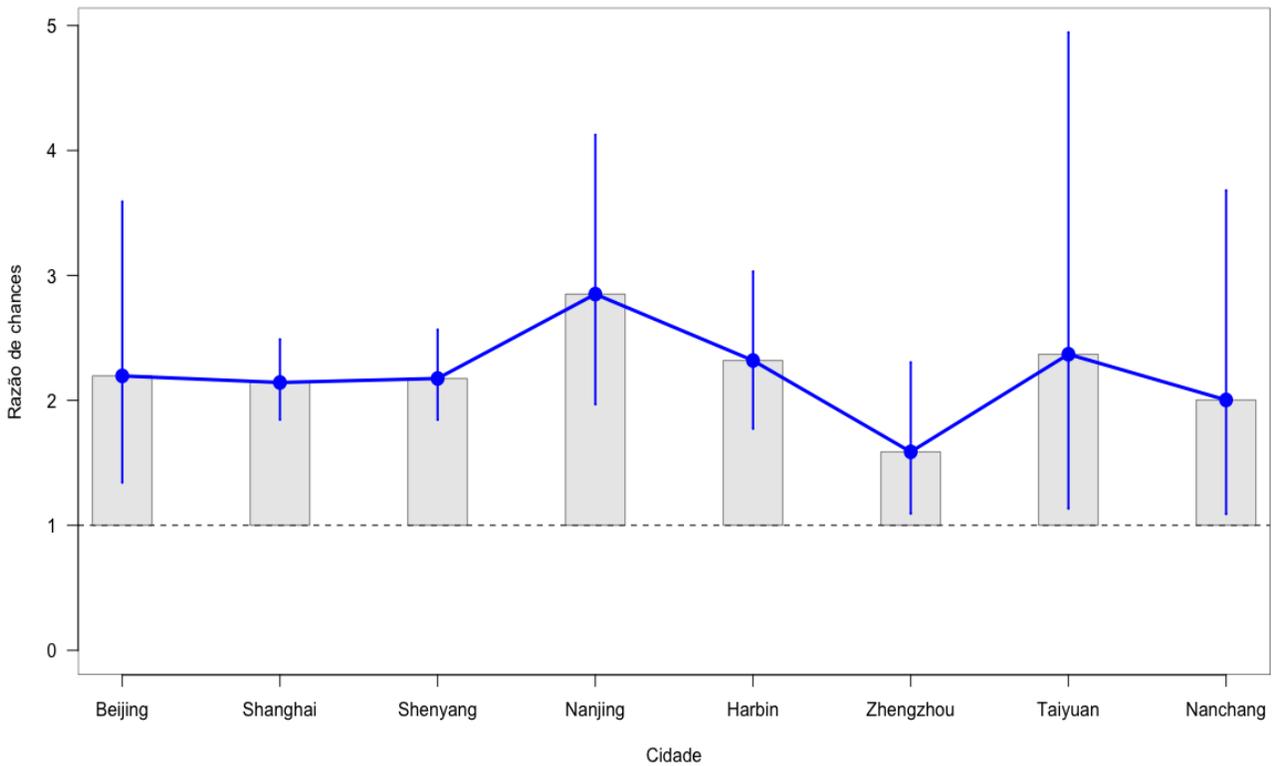
```
confint(rcparcial, level = 0.95)

      2,5 %   97,5 %
Beijing  1,343228 3,590169
Shanghai 1,845675 2,488135
Shenyang  1,844495 2,565350
Nanjing   1,969466 4,125200
Harbin    1,774534 3,030907
Zhengzhou 1,094816 2,303241
Taiyuan   1,135122 4,944715
Nanchang  1,090563 3,679615
```

Nota 2. Escreva sua interpretação das estimativas (pontual e intervalar) das razões de chances.

Uma representação gráfica das estimativas pontuais e intervalares das razões de chances condicionais é obtida com a função `plot`. O gráfico é apresentado na página seguinte.

```
plot(rcparcial, xlab = "Cidade", main = "", col = "blue",
      ylab = "Razão de chances")
```



Nota 3. Como explicar as diferenças na precisão das estimativas intervalares no gráfico acima?

Nota 4. Refaça o gráfico acima com o eixo horizontal correspondendo a estimativas decrescentes da razão de chances.

A função genérica `summary` apresenta mais resultados, incluindo as estimativas da página anterior.

```
summary(rcparcial)
```

```
z test of coefficients:
```

```

      Estimate Std. Error z value Pr(>|z|)
Beijing      2,19600    0,55076   3,9872 6,685e-05 ***
Shanghai     2,14296    0,16329 13,1238 < 2,2e-16 ***
Shenyang     2,17526    0,18306 11,8825 < 2,2e-16 ***
Nanjing      2,85034    0,53761   5,3018 1,146e-07 ***
Harbin       2,31915    0,31671   7,3225 2,433e-13 ***
Zhengzhou    1,58796    0,30129   5,2706 1,360e-07 ***
Taiyuan      2,36915    0,88940   2,6638 0,007727 **
Nanchang     2,00321    0,62147   3,2233 0,001267 **

```

```
---
```

```
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
```

Nota 5. Interprete os resultados acima.

Continuando, como as estimativas pontuais das razões de chances nas tabelas parciais não são muito discrepantes e os intervalos de confiança têm interseção, pelo menos informalmente temos alguma indicação de homogeneidade das razões de chance. Aplicaremos o teste de Cochran-Mantel-Haenszel para o teste da hipótese nula de independência condicional entre indicador de ser fumante e indicador de câncer de pulmão com a hipótese alternativa bilateral (*default* em R).

```
mantelhaen.test(dados)
```

```
Mantel-Haenszel chi-squared test with continuity correction
```

```
data: dados
Mantel-Haenszel X-squared = 279,38, df = 1, p-value < 2,2e-16
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 1,984002 2,383249
sample estimates:
common odds ratio
 2,174482
```

O valor de *CMH* é 279,38, que leva a uma forte rejeição da hipótese nula de independência condicional entre o indicador de ser fumante e o indicador de câncer de pulmão, pois o valor-*p* é da ordem de 10^{-16} . A estimativa de Mantel-Haenszel para a razão de chances comum é 2,17 e o intervalo de confiança assintótico de 95% está afastado de 1.

No resultado acima foi aplicada a correção de continuidade (*default* em R). Sem a correção de continuidade, o comando é `mantelhaen.test(dados, correct = FALSE)`.

Outras opções de hipóteses alternativas são `mantelhaen.test(dados, alternative = "greater")` e `mantelhaen.test(dados, alternative = "less")`.

Em tabelas $2 \times 2 \times K$, a função `mantelhaen.test` permite realizar um teste exato (baseado na distribuição de n_{11+} condicional nos totais marginais de cada tabela parcial), bastando especificar o argumento `exact = TRUE` (por *default*, `exact = FALSE`).

```
mantelhaen.test(dados, exact = TRUE)
```

```
Exact conditional test of independence in 2 x 2 x k tables
```

```
data: dados
S = 2930, p-value < 2,2e-16
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 1,981356 2,385078
sample estimates:
common odds ratio
 2,173726
```

Nos resultados acima, $S = n_{11+}$. Todos os resultados com a opção `exact = TRUE` são baseados na distribuição hipergeométrica, inclusive a estimativa para a razão de chances comum.

Nota 6. Para estes dados, a diferença entre os resultados assintótico e exato condicional é pequena. Por quê?

Nota 7. Procure refazer o exemplo em SAS utilizando PROC FREQ.

Nota 8. Existe uma generalização da estatística *CMH* para situações com uma tabela $I \times J \times K$ geral. Detalhes são apresentados em Agresti (2002), *Categorical Data Analysis*, 2nd ed., Wiley: New York.