

Função mantelhaen.test

Apresentamos um exemplo com a função `mantelhaen.test` do pacote `stats` em R. Os resultados são destacados em cor azul.

Os dados do exemplo podem ser encontrados nos livros de dados categorizados de A. Agresti. São 174 casos de poliomielite registrados na cidade de Des Moines em Iowa. Estudamos a associação entre ocorrência de paralisia e aplicação da vacina Salk controlando pela faixa etária (Z).

Os dados são organizados em um arranjo (*array*) tridimensional, sendo que a variável de controle (Z : idade) deve ocupar a terceira dimensão.

```
datasalk <- array(c(20, 10, 14, 24,
                  15, 3, 12, 15,
                  3, 3, 2, 2,
                  7, 1, 4, 6,
                  12, 7, 3, 5,
                  1, 3, 0, 2),
                dim = c(2, 2, 6), dimnames = list(vacina.salk = c("Sim", "Não"),
                paralisia = c("Não", "Sim"),
                idade = c("0-4", "5-9", "10-14", "15-19", "20-39", "40-")))

datasalk
, , idade = 0-4
  paralisia
vacina.salk Não Sim
           Sim  20  14
           Não  10  24
, , idade = 5-9
  paralisia
vacina.salk Não Sim
           Sim  15  12
           Não   3  15
, , idade = 10-14
  paralisia
vacina.salk Não Sim
           Sim   3   2
           Não   3   2
, , idade = 15-19
  paralisia
vacina.salk Não Sim
           Sim   7   4
           Não   1   6
, , idade = 20-39
  paralisia
vacina.salk Não Sim
           Sim  12   3
           Não   7   5
, , idade = 40-
  paralisia
vacina.salk Não Sim
           Sim   1   0
           Não   3   2
```

Cada linha na chamada da função `array` corresponde a uma tabela parcial, que é preenchida por colunas.

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`.

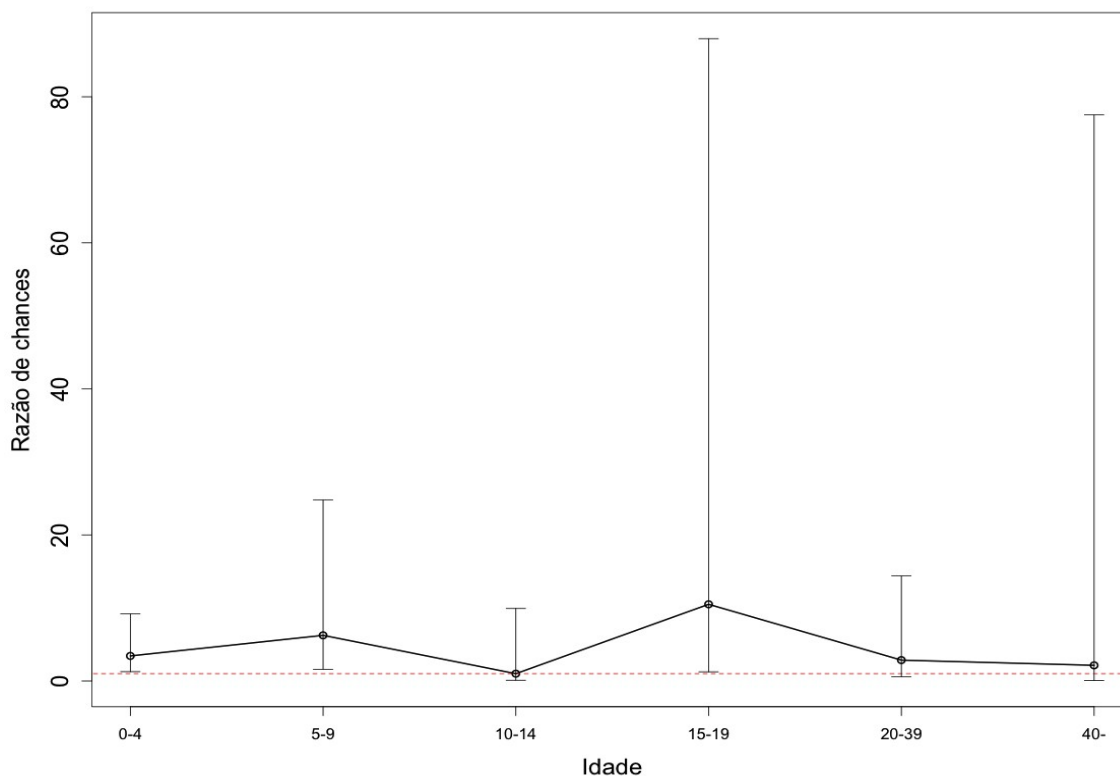
```
(rcparcial = oddsratio(datasalk, log = FALSE, stratum = 3))
      0-4      5-9      10-14      15-19      20-39      40-
3.428571  6.250000  1.000000 10.500000  2.857143  2.142857
```

Para estes dados, cada contagem é adicionada de 0,5, pois uma das contagens é igual a 0. Intervalos de confiança assintóticos de 95% são obtidos com a função `confint`.

```
confint(rcparcial)
      lwr      upr
0-4  1.27729056  9.203154
5-9  1.57456281 24.808474
10-14 0.10073371  9.927164
15-19 1.25339925 87.960799
20-39 0.56697834 14.397843
40-   0.05922227 77.535643
```

Uma representação gráfica das estimativas pontuais e intervalares das razões de chances condicionais é obtida com a função `plot`.

```
plot(rcparcial, xlab = "Idade", ylab = "Razão de chances", lwd = 2, cex.lab = 1.5,
      cex.axis = 1.5)
```



Observação 1.

1.1. Escreva sua interpretação das estimativas das razões de chances.

1.2. Estimativas dos logaritmos das razões de chances são obtidas especificando o argumento `log = TRUE`.

```
(rclog = oddsratio(datasalk, log = TRUE, stratum = 3))
      0-4      5-9      10-14      15-19      20-39      40-
1.232144e+00 1.832581e+00 -2.220446e-16 2.351375e+00 1.049822e+00 7.621401e-01
```

A função genérica `summary` apresenta mais resultados, incluindo as estimativas acima.

```
summary(rclog)
      Log Odds Ratio Std. Error z value Pr(>|z|)
0-4      1.2321e+00 5.0379e-01 2.4458 0.007227 **
5-9      1.8326e+00 7.0338e-01 2.6054 0.004589 **
10-14    -2.2204e-16 1.1711e+00 0.0000 0.500000
15-19    2.3514e+00 1.0845e+00 2.1682 0.015071 *
20-39    1.0498e+00 8.2515e-01 1.2723 0.101636
40-      7.6214e-01 1.8310e+00 0.4163 0.338612
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As estimativas dos erros padrão (em escala log) podem ser calculadas manualmente com a função `apply` aplicada à terceira dimensão do arranjo `datasalk` acrescido de 0,5. No comando abaixo, `tabela` é um argumento mudo representando cada tabela parcial (2 x 2).

```
apply(datasalk + 0.5, 3, function(tabela) sqrt(sum(1 / tabela)))
      0-4      5-9      10-14      15-19      20-39      40-
0.5037861 0.7033822 1.1710801 1.0844669 0.8251459 1.8309508
```

Como todas as estimativas pontuais das razões de chances nas tabelas parciais são maiores do que ou iguais a 1, aplicaremos o teste de Cochran-Mantel-Haenszel para o teste da hipótese nula de independência condicional (ou seja, dado Z : idade) entre `vacina.salk` e paralisia com a hipótese alternativa bilateral.

```
mantelhaen.test(datasalk)
      Mantel-Haenszel chi-squared test with continuity correction
data:  datasalk
Mantel-Haenszel X-squared = 15.3041, df = 1,
p-value = 9.152e-05
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
2.013768 7.691144
sample estimates:
common odds ratio
3.935503
```

Para estes dados o valor de CMH é 15,304. Estes resultados levam a uma forte rejeição da hipótese nula de independência condicional entre vacina.salk e paralisia a um nível de significância de 5%, pois $p < 0,0001$. A estimativa de Mantel-Haenszel para a razão de chances comum é 3,936 e o intervalo de confiança assintótico de 95% está afastado de 1.

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

Sem a correção de continuidade o comando é

```
mantelhaen.test(data26, correct = FALSE)
Mantel-Haenszel chi-squared test without
continuity correction
data:  datasalk
Mantel-Haenszel X-squared = 16.5439, df = 1,
p-value = 4.754e-05
```

Em tabelas $2 \times 2 \times K$ a função `mantelhaen.test` permite realizar um teste exato (condicional), bastando especificar o argumento `exact`.

```
mantelhaen.test(datasalk, exact = TRUE)
Exact conditional test of independence in 2 x 2 x k tables
data:  datasalk
S = 58, p-value = 4.845e-05
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 1.888774  7.871603
sample estimates:
common odds ratio
      3.795132
```

Todos os resultados com a opção `exact = TRUE` são baseados na distribuição hipergeométrica. Em particular, a estimativa para a razão de chances comum não é a de Mantel-Haenszel.

Observação 2.

2.1. Com a PROC FREQ em SAS efetue o teste de homogeneidade das razões de chances condicionais.

2.2. 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.). New York: Wiley.