

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 no livro *Categorical Data Analysis* (Agresti, A., 1990, Wiley: New York, p. 256). São 174 casos de poliomielite registrados na cidade de Des Moines em Iowa, EUA. 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. O argumento `dim = c(2, 2, 6)` especifica que a tabela é $2 \times 2 \times 6$.

```
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 (*default* em R).

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(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 (por *default*, `correct = TRUE`). O argumento `log = FALSE` indica intervalos para a razão de chances (por *default*, `log = TRUE` 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.

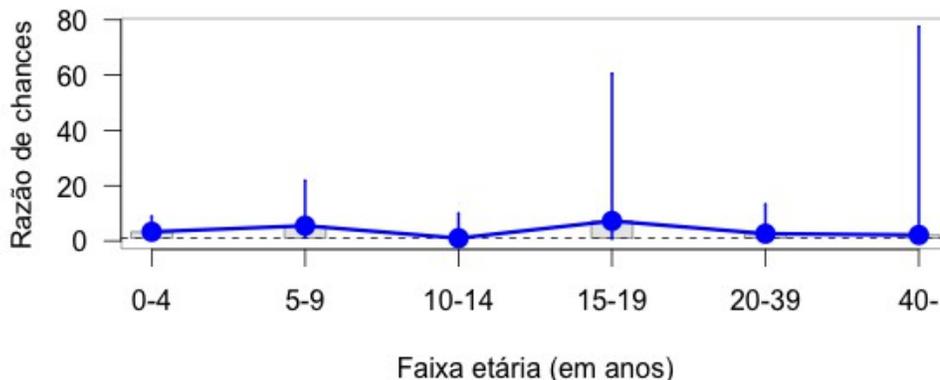
Intervalos de confiança assintóticos de 95% são obtidos com a função `confint`.

```
confint(rcparcial, level = 0.95)
```

```
      2.5 %      97.5 %
0-4  1.22896395  8.854951
5-9  1.38345587 21.797434
10-14 0.10073371  9.927164
15-19 0.86212647 60.502137
20-39 0.51973015 13.198023
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 = "Faixa etária (em anos)", main = "", col = "blue",
      ylab = "Razão de chances")
```



Nota 1. Escreva sua interpretação das estimativas das razões de chances.

Nota 2. Estimativas dos logaritmos das razões de chances são obtidas especificando o argumento `log = TRUE`.

```
(rclog = oddsratio(datasalk, log = TRUE, stratum = 3))
```

```
log odds ratios for vacina.salk and paralisia by idade
```

```
      0-4      5-9      10-14      15-19      20-39      40-
1.1935741 1.7031884 0.0000000 1.9771627 0.9628107 0.7621401
```

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

```
summary(rclog)
```

```
      Estimate Std. Error z value Pr(>|z|)
0-4      1.19357    0.50379  2.3692 0.01783 *
5-9      1.70319    0.70338  2.4214 0.01546 *
10-14    0.00000    1.17108  0.0000 1.00000
15-19    1.97716    1.08447  1.8232 0.06828 .
20-39    0.96281    0.82515  1.1668 0.24328
40-      0.76214    1.83095  0.4163 0.67722
```

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 (2×2).

```
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
```

De outra forma, `apply(datasalk, 3, function(tabela) sqrt(sum(1 / (tabela + 0.5))))`.

Continuando, 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 entre `vacina.salk` e `paralisia` com a hipótese alternativa bilateral (*default* em R).

```
mantelhaen.test(datasalk)
```

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

```
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")`.

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

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(datasalk, exact = TRUE)
```

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

```
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
```

Nos resultados acima, $S = n_{11+}$. 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.

Nota 3. Procure refazer o exemplo em SAS utilizando `PROC FREQ`.

Nota 4. 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.