

Análise de Dados Categorizados

Prática 2: Tabelas de Contingência $r \times s$

Prof. Dr. Márcio Augusto

Iniciamos carregando os pacotes necessários para essa aula prática.

```
# Pacotes necessários
library(ggplot2)
library(stargazer)
library(DescTools)
#library(gtsummary)
library(epitools)
library(vcd) # calcular as oddsratios locais
```

Exercício 1

Um estudo desenvolvido pela Profa. Denise Gonçalves, do Departamento de Otorrinolaringologia da UFMG, teve como interesse a ocorrência de manifestações otorrinolaringológicas em pacientes HIV positivos. Neste estudo, 112 pacientes foram acompanhados no período de março de 1993 a fevereiro de 1995, sendo 91 HIV positivo e 21 HIV negativo. A classificação quanto à infecção pelo HIV seguiu os critérios Do *Center for Disease Control* (CDC, 1987), sendo ela: HIV soronegativo (não possui o HIV), HIV soropositivo assintomático (possui o vírus mas não desenvolveu o quadro clínico de AIDS), com ARC (*Aids Related Complex*: apresenta baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS), ou AIDS (apresenta infecções oportunistas que definem AIDS).

As Covariáveis medidas no estudo são:

- id : Idade do Paciente (medida em anos);
- sex: Sexo do Paciente (0 se Masculino e 1 se Feminino)
- grp: Grupo de Risco (1 se HIV Soronegativo, 2 se HIV Soropositivo Assintomático, 3 se ARC e 4 se AIDS)
- ats: Atividade Sexual (1 se Homossexual, 2 se Bissexual e 3 se Heterossexual)
- ud: Uso de Droga Injetável (1 se Sim e 2 se Não)
- ac: Uso de Cocaína por Aspiração (1 se Sim e 2 se Não)

```
# lendo o banco de dados
dados<-read.table("aids.txt",h=T)
```

Vamos utilizar as variáveis Sexo e Grupo de Risco

```
#transformando as variáveis em factor e recodificando seus níveis
dados$sex <- factor(dados$sex, #vetor com os valores a serem rotulados
                    levels = c(0,1), #os valores distintos que aparecem
                    labels = c("Mas", "Fem")) #os rótulos dos valores
```

```
dados$grp <- factor(dados$grp, levels = c(1,2,3,4),
                    labels = c("HIV-", "HIV+", "ARC", "AIDS"))
```

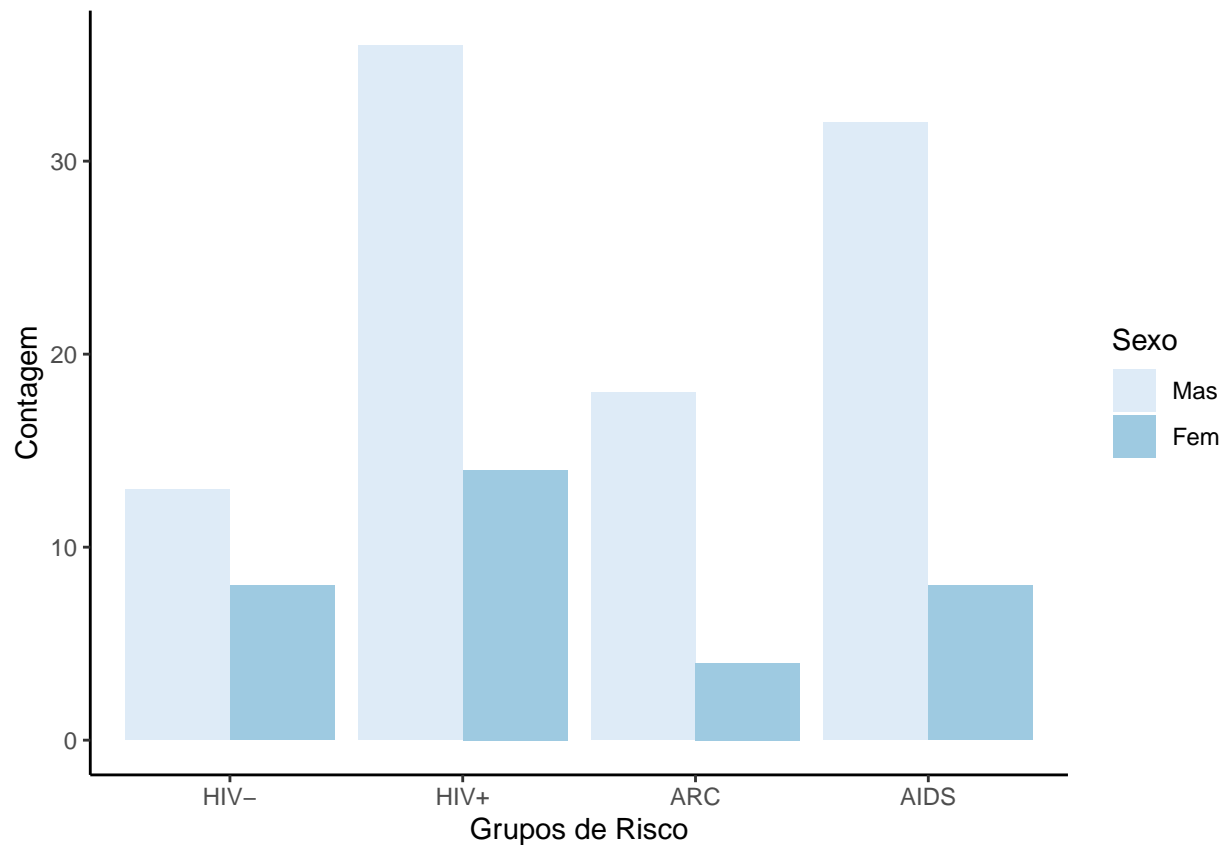
Tabelas de distribuição frequência

```
tab <- descr::CrossTable(
  x = dados$sex,
  y = dados$grp,
  dnn = c("Sexo", "Grupos de Risco"),
  prop.r = FALSE,
  prop.c = FALSE,
  prop.t = FALSE,
  prop.chisq = FALSE,
  expected = TRUE
)
tab
```

```
##      Conteúdo das células
## |-----|
## |                      N |
## |          N esperado |
## |-----|
##
## =====
##      Grupos de Risco
## Sexo   HIV-   HIV+   ARC   AIDS   Total
## -----
## Mas      13    36    18    32     99
##          15.6  37.2  16.4  29.8
## -----
## Fem       8     14     4     8     34
##          5.4   12.8   5.6   10.2
## -----
## Total    21    50    22    40    133
## =====
```

Representação gráfica

```
#grafico
ggplot(dados, aes(x=grp, fill= sex))+
  geom_bar(position = "dodge") + #especificando o formato do gráfico
  scale_fill_brewer()+
  labs(x = "Grupos de Risco", #rótulo do eixo x
       y = "Contagem", #rótulo do eixo y
       fill = "Sexo") + #rótulo da legenda
  theme_classic() #modificando o tema do gráfico
```



Vamos realizar o teste de Qui-quadrado

```
chisq.test(dados$sex,dados$grp, correct = FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data: dados$sex and dados$grp
## X-squared = 3.1697, df = 3, p-value = 0.3662
```

Cálculo das razões de chances nominais

```
### Cálculo das Odds ratios nominais
```

```
tab<-table(dados$sex,dados$grp)
```

```
tab[,c(1,4)]
```

```
##
##      HIV- AIDS
##  Mas   13   32
##  Fem    8    8
```

```
ods14<- epitools::oddsratio(tab[,c(1,4)])$measure[-1,]
ods14
```

```
## estimate      lower      upper
```

```
## 0.4136053 0.1229002 1.3691107
```

```
ods14<- epitools::oddsratio(tab[,c(1,4)])$measure[-1,]  
ods14
```

```
## [1] 0.4136053
```

```
1/ods14
```

```
## [1] 2.417764
```

Assim, a chance de ocorrência de uma pessoa do sexo masculino com AIDS é, aproximadamente, 2,5 vezes a chance da ocorrência de uma pessoa do sexo masculino HIV negativo.

```
tab[,c(2,4)]
```

```
##  
##      HIV+ AIDS  
##  Mas   36   32  
##  Fem   14    8
```

```
ods24<- epitools::oddsratio(tab[,c(2,4)])$measure[-1,]  
ods24
```

```
## estimate      lower      upper  
## 0.6501084 0.2292865 1.7412750
```

```
ods24<- epitools::oddsratio(tab[,c(2,4)])$measure[-1,]  
ods24
```

```
## [1] 0.6501084
```

```
1/ods24
```

```
## [1] 1.538205
```

Assim, a chance de ocorrência de uma pessoa do sexo masculino com AIDS é, aproximadamente, 1,5 vezes a chance da ocorrência de uma pessoa do sexo masculino HIV positivo.

```
tab[,c(3,4)]
```

```
##  
##      ARC AIDS  
##  Mas   18   32  
##  Fem    4    8
```

```
ods34<- epitools::oddsratio(tab[,c(3,4)])$measure[-1,]
```

```
## Warning in chisq.test(xx, correct = correction): Aproximação do qui-quadrado  
## pode estar incorreta
```

```
ods34
```

```
## estimate      lower      upper  
## 1.1065967 0.2952903 4.8085351
```

```
ods34<- epitools::oddsratio(tab[,c(3,4)])$measure[-1,]
```

```
## Warning in chisq.test(xx, correct = correction): Aproximação do qui-quadrado  
## pode estar incorreta
```

```
ods34
```

```
## [1] 1.106597
```

Assim, a chance de ocorrência de uma pessoa do sexo masculino com ARC é, aproximadamente, 10% maior do que chance da ocorrência de uma pessoa do sexo masculino com AIDS.

Exercício 2

A reação ao tratamento por quimioterapia está sendo estudada em quatro grupos de pacientes com câncer. Deseja-se investigar se todos os tipos reagem da mesma maneira. Uma amostra de pacientes de cada grupo foi escolhida ao acaso e classificou-se a reação em três categorias:

```
dados<-matrix(c(51,33,16,58,29,13,48,42,30,26,38,16), byrow = T, ncol=3)
dimnames(dados) <- list(Cancer = c("I","II","III","IV"), Reação = c("Pouco", "Média", "Alta"))
stargazer(addmargins(dados), type = "text", title="Dados do estudo", digits = 1)
```

```
##
## Dados do estudo
## =====
##      Pouco Média Alta Sum
## -----
## I      51      33      16 100
## II     58      29      13 100
## III    48      42      30 120
## IV     26      38      16  80
## Sum   183     142      75 400
## -----
```

1 Teste Qui-quadrado de Pearson

```
chisq.test(dados, correct = F)
```

```
##
## Pearson's Chi-squared test
##
## data:  dados
## X-squared = 17.173, df = 6, p-value = 0.008669
```

2 Razão de chances local

A função loddsratio() do pacote vcd calcula o logarítmo da razão de chances.

```
log_RC <- loddsratio(dados)
log_RC
```

```
## log odds ratios for Cancer and Reação
##
##      Reação
## Cancer Pouco:Média Média:Alta
## I:II    -0.2578291 -0.07842763
## II:III   0.5596158  0.46587424
## III:IV   0.5130210 -0.52852520
```

```
confint(log_RC)
```

```
##              2.5 %    97.5 %
## I:II/Pouco:Média -0.8826705 0.3670122
## II:III/Pouco:Média -0.0488165 1.1680481
```

```
## III:IV/Pouco:Média -0.1353100 1.1613521
## I:II/Média:Alta -0.9641251 0.8072698
## II:III/Média:Alta -0.3387827 1.2705312
## III:IV/Média:Alta -1.2773203 0.2202699

coef(log_RC)

## I:II/Pouco:Média II:III/Pouco:Média III:IV/Pouco:Média I:II/Média:Alta
## -0.25782911 0.55961579 0.51302101 -0.07842763
## II:III/Média:Alta III:IV/Média:Alta
## 0.46587424 -0.52852520

summary(log_RC)

##
## z test of coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## I:II/Pouco:Média -0.257829 0.318802 -0.8087 0.41866
## II:III/Pouco:Média 0.559616 0.310430 1.8027 0.07143 .
## III:IV/Pouco:Média 0.513021 0.330787 1.5509 0.12092
## I:II/Média:Alta -0.078428 0.451895 -0.1736 0.86222
## II:III/Média:Alta 0.465874 0.410547 1.1348 0.25647
## III:IV/Média:Alta -0.528525 0.382045 -1.3834 0.16654
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Calculando as razões de chances locais
RC_local<-exp(coef(log_RC))
RC_local

## I:II/Pouco:Média II:III/Pouco:Média III:IV/Pouco:Média I:II/Média:Alta
## 0.7727273 1.7500000 1.6703297 0.9245690
## II:III/Média:Alta III:IV/Média:Alta
## 1.5934066 0.5894737

## Calculando os intervalos de confiança
exp(confint(log_RC))

## 2.5 % 97.5 %
## I:II/Pouco:Média 0.4136767 1.443416
## II:III/Pouco:Média 0.9523559 3.215710
## III:IV/Pouco:Média 0.8734451 3.194249
## I:II/Média:Alta 0.3813167 2.241779
## II:III/Média:Alta 0.7126373 3.562745
## III:IV/Média:Alta 0.2787834 1.246413
```

Exercício 3

Os dados a seguir são de um estudo apresentado por Glassman et al. (1990), que teve como objetivo avaliar o Hábito de fumar vs. transtorno depressivo grave por gênero para uma amostra coletada no centro epidemiológico de St. Louis.

```
freq <- c(40, 10, 889, 417, 104, 40, 840, 873)
names <- list(Fumo=c("Sim", "Nao"), Depressao=c("Sim", "Nao"),
```

```

Sexo=c("Homem", "Mulher")
dados <- array(freq, c(2,2,2), dimnames=names)
dados

```

```

## , , Sexo = Homem
##
##      Depressao
## Fumo  Sim Nao
##      Sim  40 889
##      Nao  10 417
##
## , , Sexo = Mulher
##
##      Depressao
## Fumo  Sim Nao
##      Sim 104 840
##      Nao  40 873

```

```

# A função ftable() pode ser utilizada para imprimir uma versão mais
# compacta e atraente da tabela.
ftable(dados)

```

```

##              Sexo Homem Mulher
## Fumo Depressao
## Sim  Sim           40      104
##      Nao           889      840
## Nao  Sim           10       40
##      Nao           417      873

```

Tabelas Parciais

```

## Tabelas Parciais
dados[, ,1] #homens

```

```

##      Depressao
## Fumo  Sim Nao
##      Sim  40 889
##      Nao  10 417

```

```

dados[, ,2] #mulheres

```

```

##      Depressao
## Fumo  Sim Nao
##      Sim 104 840
##      Nao  40 873

```

Tabelas Marginais

```

## Tabelas marginais
margin.table(dados, c(1,2))

```

```
##      Depressao
## Fumo  Sim  Nao
##      Sim 144 1729
##      Nao  50 1290
```

```
dados[1,,]
```

```
##      Sexo
## Depressao Homem Mulher
##      Sim    40    104
##      Nao   889    840
```

```
****
```

Odds ratio partial

```
## Odds ratio partial
epitools::oddsratio(dados[, ,1]) #Homens
```

```
## $data
##      Depressao
## Fumo  Sim  Nao Total
##      Sim   40  889  929
##      Nao   10  417  427
##      Total  50 1306 1356
##
## $measure
##      odds ratio with 95% C.I.
## Fumo estimate      lower      upper
##      Sim 1.000000      NA      NA
##      Nao 1.853411 0.9521231 3.976592
##
## $p.value
##      two-sided
## Fumo midp.exact fisher.exact chi.square
##      Sim      NA      NA      NA
##      Nao 0.07054359 0.0875684 0.0746949
##
## $correction
## [1] FALSE
##
## attr("method")
## [1] "median-unbiased estimate & mid-p exact CI"
```

```
epitools::oddsratio(dados[, ,1])$measure[-1,]
```

```
## estimate      lower      upper
## 1.8534109 0.9521231 3.9765925
```

```
epitools::oddsratio(dados[, ,2]) #Mulheres
```

```
## $data
##      Depressao
## Fumo  Sim  Nao Total
##      Sim 104  840  944
##      Nao  40  873  913
##      Total 144 1713 1857
```



```
##
## $measure
##      odds ratio with 95% C.I.
## Fumo  estimate      lower  upper
##   Sim 1.000000      NA      NA
##   Nao 2.694351 1.862663 3.97131
##
## $p.value
##      two-sided
## Fumo  midp.exact fisher.exact  chi.square
##   Sim      NA      NA      NA
##   Nao 6.16224e-08 7.993251e-08 9.034205e-08
##
## $correction
## [1] FALSE
##
## attr("method")
## [1] "median-unbiased estimate & mid-p exact CI"
epitools::oddsratio(dados[, ,2])$measure[-1,]

## estimate      lower      upper
## 2.694351 1.862663 3.971310
```

Teste de Mantel-Haenszel

```
## Teste de Mantel-Haenszel
mantelhaen.test(dados, correct = FALSE)

##
## Mantel-Haenszel chi-squared test without continuity correction
##
## data: dados
## Mantel-Haenszel X-squared = 30.618, df = 1, p-value = 3.141e-08
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.779109 3.463946
## sample estimates:
## common odds ratio
##      2.482486
```