

Listas V, VI e VII de Exercícios - Análise de Dados Categóricos

Bruna Wundervald

Novembro de 2016

Lista V

1. Os dados exibidos na **Tabela 1** são de um estudo sobre doença coronária (CHD) em que CAT = nível de catecholamine (0 se baixo e 1 se alto), IDADE (0 se < 55 e 1 se = 55 anos) e ECG = eletrocardiograma (0 se normal e 1 se anormal).
- (a) Ajuste um modelo de regressão logística aos dados desse estudo e apresente conclusões. Avalie o efeito das interações duplas

```
library(xtable)
da1 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/ex1chd.txt",
                  header = TRUE)

# da1 <- read.table("da.txt")
print(xtable(da1, type = "html", caption = "Doença Coronária"))
```

% latex table generated in R 3.3.1 by xtable 1.8-2 package % Wed Nov 16 17:47:46 2016

	sim	nao	cat	idade	ecg
1	17	257	0	0	0
2	15	107	0	1	0
3	7	52	0	0	1
4	5	27	0	1	1
5	1	7	1	0	0
6	9	30	1	1	0
7	3	14	1	0	1
8	14	44	1	1	1

Tabela 1: Doença Coronária

```
m0 <- glm(as.matrix(da1[,c(1,2)]) ~ (cat+idade+ecg)^2,
          family=binomial(link="logit"),
          data=da1)
anova(m0, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: as.matrix(da1[, c(1, 2)])
##
## Terms added sequentially (first to last)
```

```
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      21.3320
## cat           1  14.1312           6       7.2008 0.0001705 ***
## idade         1   4.7240           5       2.4768 0.0297448 *
## ecg           1   1.5224           4       0.9544 0.2172564
## cat:idade     1   0.0321           3       0.9223 0.8577925
## cat:ecg       1   0.5035           2       0.4188 0.4779635
## idade:ecg     1   0.4157           1       0.0031 0.5190999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como observado acima, as variáveis idade e nível de catecholamine são significativas para o modelo. Podemos comentar que as interações duplas não são significativas.

A *deviance* do modelo ajustado é, o que é um resultado bom quando precisamos avaliar a qualidade do ajuste. Quanto menor é esta *deviance*, melhor o nosso modelo está declarado.

Em um segundo ajuste, mantemos as duas variáveis anteriormente relevantes de acordo com a anova do modelo:

```
m0 <- glm(as.matrix(da1[,c(1,2)]) ~ cat+idade,
          family = binomial(link = "logit"),
          data = da1)
summary(m0)

##
## Call:
## glm(formula = as.matrix(da1[, c(1, 2)]) ~ cat + idade, family = binomial(link = "logit"),
##      data = da1)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7
## -0.72188 -0.15477  1.24039  0.47158 -0.17198 -0.14599  0.34611
##      8
##  0.01182
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.5397     0.2005 -12.664 < 2e-16 ***
## cat           0.7736     0.2955   2.618  0.00884 **
## idade        0.6174     0.2840   2.173  0.02975 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 21.3320  on 7  degrees of freedom
## Residual deviance:  2.4768  on 5  degrees of freedom
## AIC: 37.188
##
## Number of Fisher Scoring iterations: 4
```

Aqui, obtemos um modelo para o qual ambas as variáveis são consideravelmente significativas.

```
dev <-residuals(m0, type='deviance')
QL <- sum(dev^2)
p1 <- 1-pchisq(QL, 5)
cbind(QL,p1)
```

```
##           QL           p1
## [1,] 2.47683 0.7799804
```

```
rpears <- residuals(m0, type='pearson')
QP <- sum(rpears^2)
p2 <-1-pchisq(QP, 5)
cbind(QP, p2)
```

```
##           QP           p2
## [1,] 2.735735 0.7406478
```

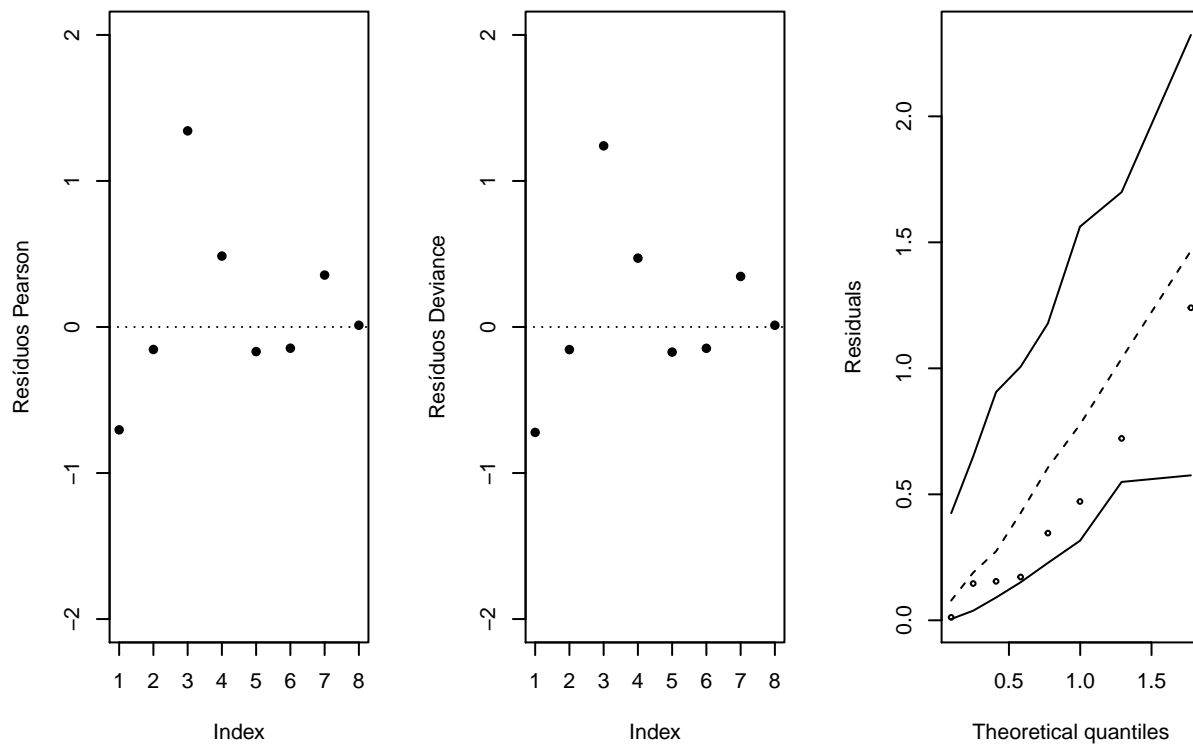
```
par(mfrow=c(1,3))
plot(rpears, ylab="Resíduos Pearson", pch=16, ylim=c(-2,2))
abline(h=0, lty=3)
plot(dev, ylab="Resíduos Deviance", pch=16, ylim=c(-2,2))
abline(h=0, lty=3)

#-----
library(hnp)
```

```
## Loading required package: MASS
```

```
hnp(m0)
```

```
## Binomial model
```



2. Um estudo reuniu informações entre 1994 e 1995 de 494 indivíduos que sofreram acidente traumático e foram atendidos pelo SIATE (Serviço Integrado de Atendimento ao Trauma em Emergência). A fim de prever a probabilidade de óbito nas primeiras 24 horas após o acidente, foi ajustado um modelo de regressão logística aos dados do estudo. O modelo ajustado ficou expresso por:

$$\ln\left(\frac{\hat{p}(x)}{1 - \hat{p}(x)}\right) = 2.211 + 2.607x_1 - 0.52x_2,$$

em que x_1 = número de lesões no tórax, podendo variar de 0 a 5, e x_2 = escala de coma de Glasgow (GCS) = total registrado para cada indivíduo no Quadro 1, que pode variar entre 3 e 15.

- (a) Estime as probabilidades $p(x)$ para todas as combinações de x_1 e x_2 apresentando-as em ordem decrescente a fim de identificar os indivíduos que necessitam de encaminhamento hospitalar com muita, moderada ou pouca urgência

```
# les <- rep(c(0:5), c(13, 13, 13, 13, 13, 13))
# coma <- rep(3:15, rep(6, 13))
# da2 <- cbind(les, coma)
#
# ## checar se faz sentido coma = 0 = referência
# model <- function(x)
# {
#   n <- nrow(x)
#   #
#   for(i in 1:n){
#     x1 <- x[,1]
#     x2 <- x[,2]
#     m <- 2.211 + 2.607 * x1 - 0.52 * x2
```

```

#   resp <- exp(m)/(1+exp(m))
#   }
#   return(resp)
#   }
# n <- model(da2)
# cbind(da2, n)
# hist(n)

x1<-rep(0:5,13)
x2<-rep(3:15,rep(6,13))
px<-(exp(2.211 + 2.607*x1 - 0.52*x2))/(1+ exp(2.211 + 2.607*x1 - 0.52*x2))
est<-as.data.frame(cbind(x1,x2,px))
i<-order(est[,3], decreasing=T)
px_ord<-est[i,]
px_ord

```

```

##      x1 x2      px
## 6    5  3 0.999998862
## 12   5  4 0.999998085
## 18   5  5 0.999996780
## 24   5  6 0.999994583
## 30   5  7 0.999990889
## 36   5  8 0.999984675
## 5    4  3 0.999984567
## 42   5  9 0.999974223
## 11   4  4 0.999974042
## 48   5 10 0.999956643
## 17   4  5 0.999956338
## 54   5 11 0.999927075
## 23   4  6 0.999926562
## 60   5 12 0.999877343
## 29   4  7 0.999876482
## 66   5 13 0.999793706
## 35   4  8 0.999792257
## 4    3  3 0.999790798
## 72   5 14 0.999653056
## 41   4  9 0.999650620
## 10   3  4 0.999648166
## 78   5 15 0.999416569
## 47   4 10 0.999412473
## 16   3  5 0.999408348
## 53   4 11 0.999012159
## 22   3  6 0.999005226
## 59   4 12 0.998339542
## 28   3  7 0.998327898
## 65   4 13 0.997210223
## 34   3  8 0.997190681
## 3    2  3 0.997171003
## 71   4 14 0.995316430
## 40   3  9 0.995283685
## 9    2  4 0.995250712
## 77   4 15 0.992147190
## 46   3 10 0.992092463

```

```
## 15 2 5 0.992037359
## 52 3 11 0.986770653
## 21 2 6 0.986678960
## 58 3 12 0.977946853
## 27 2 7 0.977795379
## 64 3 13 0.963455655
## 33 2 8 0.963208392
## 2 1 3 0.962959520
## 70 3 14 0.940026203
## 39 2 9 0.939630347
## 8 1 4 0.939232046
## 76 3 15 0.903086543
## 45 2 10 0.902472163
## 14 1 5 0.901854312
## 51 2 11 0.846186636
## 20 1 6 0.845273343
## 57 2 12 0.765845626
## 26 1 7 0.764588008
## 63 2 13 0.660382673
## 32 1 8 0.658810973
## 1 0 3 0.657235775
## 69 2 14 0.536186620
## 38 1 9 0.534445353
## 7 0 4 0.532703245
## 75 2 15 0.407333400
## 44 1 10 0.405644612
## 13 0 5 0.403958054
## 50 1 11 0.288639670
## 19 0 6 0.287204512
## 56 1 12 0.194348238
## 25 0 7 0.193254544
## 62 1 13 0.125428300
## 31 0 8 0.124662439
## 68 1 14 0.078565429
## 37 0 9 0.078060172
## 74 1 15 0.048245710
## 43 0 10 0.047925298
## 49 0 11 0.029057235
## 55 0 12 0.017481088
## 61 0 13 0.010467059
## 67 0 14 0.006249405
## 73 0 15 0.003724838
```

3. A fim de avaliar a toxicidade aguda de duas drogas (A e B), estas foram administradas em dosagens diferentes (injeção intravenosa) a grupos de camundongos. As mortes foram registradas após 5 minutos da administração da droga. Os dados estão na Tabela 2.

```
da2 <-read.table("https://docs.ufpr.br/~giolo/CE073/Dados/ex3bio.txt",
                 header = TRUE)
da2$ldose<-log10(da2$dose)

# da2 <- read.table("da2.txt")
print(xtable(da2, type = "html", caption = "Bioensaio
            com Camundongos"))
```

	droga	dose	sim	nao	ldose
1	A	2	2	18	0.30
2	A	4	9	11	0.60
3	A	8	14	6	0.90
4	A	16	19	1	1.20
5	B	2	1	19	0.30
6	B	4	6	14	0.60
7	B	8	14	6	0.90
8	B	16	17	3	1.20

Tabela 2: Bioensaio com Camundongos

Considerando:

$$X_1 = \begin{cases} 1, & \text{se } droga = A \\ 0, & \text{se } droga = B \end{cases}$$

e $X_2 = \ln_{10}(dose)$.

- (a) Ajuste um modelo de regressão binomial (logístico, probito, clog-log ou Cauchy) aos dados descritos e apresente conclusões.

```
m2 <-glm(as.matrix(da2[,c(3,4)]) ~
          (droga + ldose)^2,
          family = binomial(link="logit"),
          data = da2)
anova(m2, test="Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                7      75.085
## droga             1    0.901      6      74.184  0.3424
## ldose             1   72.541      5      1.643 <2e-16 ***
## droga:ldose      1    0.003      4      1.640  0.9569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m2 <-glm(as.matrix(da2[,c(3,4)]) ~ ldose,
          family = binomial(link="logit"),
          data = da2)
anova(m2, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                7      75.085
## ldose  1      71.944         6      3.141 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# função de ligação c-loglog
m3 <-glm(as.matrix(da2[,c(3,4)]) ~
        (droga + ldose)^2,
        family = binomial(link="cloglog"),
        data = da2)
anova(m3, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: cloglog
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                7      75.085
## droga      1      0.901         6      74.184  0.3424
## ldose      1     70.176         5       4.007 <2e-16 ***
## droga:ldose 1      0.000         4       4.007  0.9933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m3 <-glm(as.matrix(da2[,c(3,4)]) ~ ldose,
        family = binomial(link="cloglog"),
        data = da2)
anova(m3, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: cloglog
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
```



```
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      75.085
## ldose  1    69.364                6      5.721 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

função de ligação probito

```
m4 <-glm(as.matrix(da2[,c(3,4)]) ~
        (droga + ldose)^2,
        family = binomial(link="probit"),
        data = da2)
anova(m4, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: probit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      75.085
## droga      1     0.901                6      74.184  0.3424
## ldose      1    72.735                5       1.449 <2e-16 ***
## droga:ldose 1     0.009                4       1.440  0.9256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m4 <-glm(as.matrix(da2[,c(3,4)]) ~ ldose,
        family = binomial(link="probit"),
        data = da2)
anova(m4, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: probit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      75.085
## ldose  1    72.01                6      3.075 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

função de ligação cauchy

```
m5 <-glm(as.matrix(da2[,c(3,4)]) ~
```

```

      (droga + ldose)^2,
      family = binomial(link="cauchit"),
      data = da2)
anova(m5, test="Chisq")

```

```

## Analysis of Deviance Table
##
## Model: binomial, link: cauchit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      75.085
## droga          1      0.901          6      74.184  0.3424
## ldose          1     70.500          5       3.683 <2e-16 ***
## droga:ldose    1      0.009          4       3.674  0.9245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

m5 <-glm(as.matrix(da2[,c(3,4)]) ~ ldose,
      family = binomial(link="cauchit"),
      data = da2)
anova(m5, test="Chisq")

```

```

## Analysis of Deviance Table
##
## Model: binomial, link: cauchit
##
## Response: as.matrix(da2[, c(3, 4)])
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                7      75.085
## ldose    1      70.13          6      4.955 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Notamos que a variável “droga” não se mostrou significativa. Isto quer dizer que o nível de toxicidade aguda de cada droga não influencia nas mortes dos camundongos. Apenas o log da dose aplicada de cada droga parece ter efeito sobre os animais.

```

cbind(m2$aic, m3$aic, m4$aic, m5$aic)

```

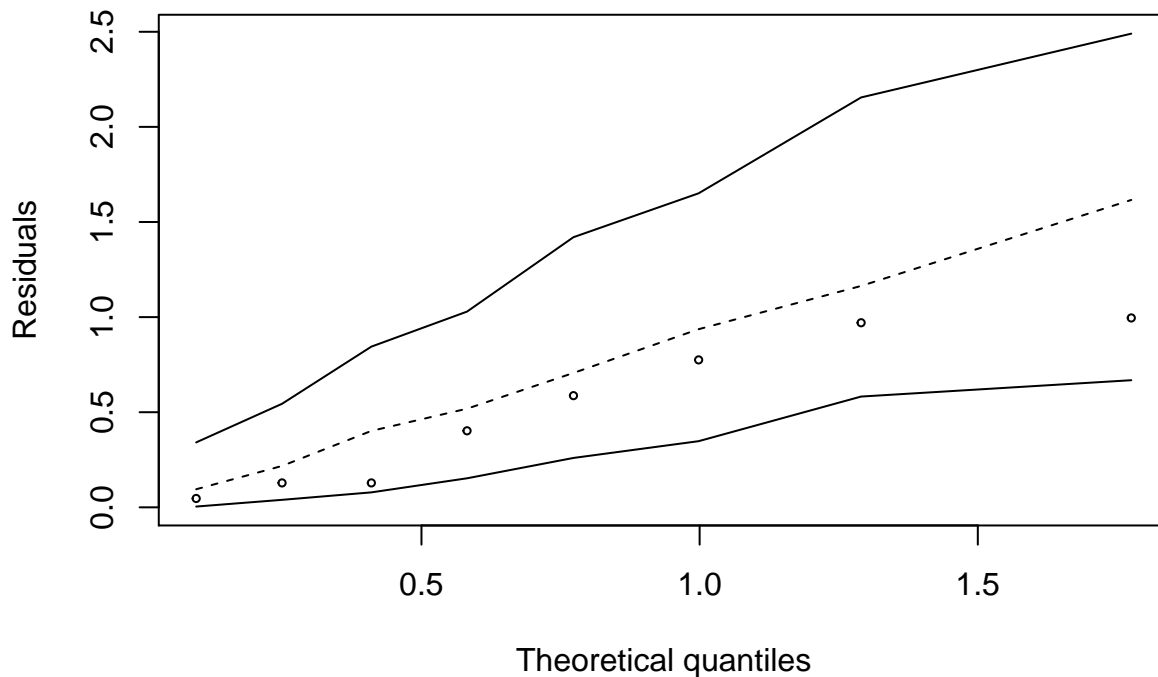
```

##          [,1]    [,2]    [,3]    [,4]
## [1,] 29.75466 32.3345 29.68889 31.56915

```

```
# os modelos com função de ligação
# probito e logito têm os menores
# AIC
par(mfrow = c(1,1))
hnp(m4)
```

```
## Binomial model
```



```
# valores preditos
f <- m4$fitted.values
OR <- (f[4]/(1-f[4]))/(f[3]/(1-f[3])); OR
```

```
## 4
## 4.984086
```

Tem-se, por exemplo, que os camundongos que receberam dosagem 16 ($\log_{10}(\text{dose}) = 1.20412$) apresentaram chance de morte igual a aproximadamente 5 vezes a daqueles que receberam dosagem 8 ($\log_{10}(\text{dose}) = 0.90309$).

4. Um grupo de 4.587 indivíduos sem doença cardíaca coronária (CHD) ao ingressarem no *Framingham Heart Study* foi acompanhado por 12 anos registrando-se, ao final desse período, os que desenvolveram a doença. Os dados por sexo, grupo de idade (em anos) e nível de colesterol inicial (mg/100ml) estão na Tabela 33.

- (a) Ajuste um modelo de regressão logística aos dados desse estudo.
- (b) Com base no modelo ajustado em (a) é possível concluir que as variáveis X_1 , X_2 e X_3 são fatores de risco para CHD?

```

# i <- expand.grid( sexo = c("Fem", "Masc"),
#                  idade = c("30-49", "50-62"),
#                  nivel = c("<190", "190-219", "220-249", "250<"),
#                  s = 0, n = 0)
# da4 <- edit(i)
# write.table(da4, file = "da4.txt")

da4 <- read.table("da4.txt")
# da4 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/ex5rote.txt",
#                  # header = TRUE)
print(xtable(da4, type = "html", caption = "Framingham
Heart Study"))

```

```

## % latex table generated in R 3.3.1 by xtable 1.8-2 package
## % Wed Nov 16 17:47:50 2016
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllllr}
## \hline
## & sexo & idade & nivel & s & n \\\
## \hline
## 1 & Fem & 30-49 & $<$190 & 6 & 536 \\\
## 2 & Masc & 30-49 & $<$190 & 13 & 327 \\\
## 3 & Fem & 50-62 & $<$190 & 9 & 49 \\\
## 4 & Masc & 50-62 & $<$190 & 13 & 110 \\\
## 5 & Fem & 30-49 & 190-219 & 5 & 547 \\\
## 6 & Masc & 30-49 & 190-219 & 18 & 390 \\\
## 7 & Fem & 50-62 & 190-219 & 12 & 123 \\\
## 8 & Masc & 50-62 & 190-219 & 33 & 143 \\\
## 9 & Fem & 30-49 & 220-249 & 10 & 402 \\\
## 10 & Masc & 30-49 & 220-249 & 40 & 381 \\\
## 11 & Fem & 50-62 & 220-249 & 21 & 197 \\\
## 12 & Masc & 50-62 & 220-249 & 35 & 139 \\\
## 13 & Fem & 30-49 & 250$<$ & 18 & 339 \\\
## 14 & Masc & 30-49 & 250$<$ & 57 & 305 \\\
## 15 & Fem & 50-62 & 250$<$ & 48 & 347 \\\
## 16 & Masc & 50-62 & 250$<$ & 49 & 134 \\\
## \hline
## \end{tabular}
## \caption{Framingham
##          Heart Study}
## \end{table}

```

```

m4 <- glm(cbind(s, n)~ . , data = da4,
          family = binomial)
summary(m4)

```

```

##
## Call:
## glm(formula = cbind(s, n) ~ ., family = binomial, data = da4)
##
## Deviance Residuals:

```

```
##      Min      1Q   Median      3Q      Max
## -1.9244 -0.7387 -0.5549  0.8537  3.1917
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.1831     0.1902 -21.993 < 2e-16 ***
## sexoMasc       1.1000     0.1162   9.467 < 2e-16 ***
## idade50-62     1.1345     0.1113  10.195 < 2e-16 ***
## nivel190-219   0.2462     0.2059   1.196  0.23176
## nivel220-249   0.7040     0.1928   3.652  0.00026 ***
## nivel250<      1.1614     0.1843   6.301 2.96e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 301.215  on 15  degrees of freedom
## Residual deviance:  26.582  on 10  degrees of freedom
## AIC: 113.26
##
## Number of Fisher Scoring iterations: 4
```

```
anova(m4, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(s, n)
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                15    301.215
## sexo  1    79.710             14    221.504 < 2.2e-16 ***
## idade 1   133.124             13     88.381 < 2.2e-16 ***
## nivel 3    61.798             10     26.582 2.426e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Aqui, conclui-se que ser do sexo masculino incorre numa maior probabilidade de ter doença coronária cardíaca, assim como estar na faixa entre 50 e 62 anos. Porém, podemos notar que não há diferença se o indivíduo tem nível de colesterol inicial abaixo de 190 ou entre 190 e 219. Esta variável só passa a ser significativa a partir do terceiro nível, ou seja, entre 220 e 249.

A chance de um indivíduo do sexo masculino ter a doença é aproximadamente 3 vezes a do sexo feminino, mantendo as classes das outras variáveis fixas.

```
exp(m4$coefficients[2])
```

```
## sexoMasc
## 3.00405
```

5. Para avaliar a toxicidade do inseticida rotenone, um bioensaio foi conduzido sob o delineamento completamente casualizado. Doses crescentes do inseticida foram aplicadas a grupos de insetos (*macro-siphoniella sanborni*) registrando-se, após certo tempo, o número de insetos mortos em cada grupo. Os resultados estão na tabela abaixo.

```
# i <- expand.grid(dose = 0,
#                 s = 0, n = 0, tot = 0)
# da5 <- edit(i)
# str(da5)
# write.table(da5, file = "da5.txt")

da5 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/ex5rote.txt",
                  header = TRUE)

# da5 <- read.table("da5.txt")
print(xtable(da5, type = "html", caption = ""))
```

```
## % latex table generated in R 3.3.1 by xtable 1.8-2 package
## % Wed Nov 16 17:47:50 2016
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrr}
## \hline
## & dose & sim & nao \\
## \hline
## 1 & 0.00 & 0 & 49 \\
## 2 & 2.60 & 6 & 44 \\
## 3 & 3.80 & 16 & 32 \\
## 4 & 5.10 & 24 & 22 \\
## 5 & 7.70 & 42 & 7 \\
## 6 & 10.20 & 44 & 6 \\
## \hline
## \end{tabular}
## \caption{}
## \end{table}
```

- (a) Para os dados do bioensaio descrito, ajuste um modelo de regressão dentre os que foram discutidos no capítulo.
- (b) A partir do modelo ajustado obtenha as doses letais 50% e 90% denotadas, respectivamente, por DL_{50} e DL_{90} .

```
da5$dose[1] <- 0.01
attach(da5)
# função de ligação logito
m5 <- glm(cbind(sim, nao) ~ log10(dose), data = da5,
          family = binomial(link="logit"))
summary(m5)
```

```
##
## Call:
## glm(formula = cbind(sim, nao) ~ log10(dose), family = binomial(link = "logit"),
##      data = da5)
```

```
##
## Deviance Residuals:
##      1      2      3      4      5      6
## -0.00068 -0.16460  0.16456 -0.27929  0.87648 -0.72364
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.8869      0.6429  -7.601 2.94e-14 ***
## log10(dose)   7.1462      0.8928   8.005 1.20e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 163.7446  on 5  degrees of freedom
## Residual deviance:  1.4241  on 4  degrees of freedom
## AIC: 24.644
##
## Number of Fisher Scoring iterations: 6
```

```
x<-seq(-2,1.5,0.1)
m1<-exp(-4.8869+7.1462*x)/(1+exp(-4.8869+7.1462*x))
par(mfrow=c(1,1))

# plot(log(dose), sim/(sim+nao),
#       pch = 16,
#       ylab = "proporção de mortes",
#       xlab = "log10(dose)",
#       xlim = c(-2.1,1.5),
#       ylim = c(0,1.05))
# lines(x, m5, lty=1,lwd=2, col=2)

# função de ligação probito
m5.1 <- glm(cbind(sim, nao) ~ log10(dose), data = da5,
            family = binomial(link="probit"))
summary(m5.1)
```

```
##
## Call:
## glm(formula = cbind(sim, nao) ~ log10(dose), family = binomial(link = "probit"),
##      data = da5)
##
## Deviance Residuals:
##      1      2      3      4      5      6
##  0.00000 -0.15682  0.07474 -0.21171  1.01391 -0.79747
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.8875      0.3510  -8.225 <2e-16 ***
## log10(dose)   4.2132      0.4806   8.767 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 163.745 on 5 degrees of freedom
## Residual deviance: 1.739 on 4 degrees of freedom
## AIC: 24.959
##
## Number of Fisher Scoring iterations: 6
```

```
# plot(log(dose), sim/(sim+nao),
# col='red',pch=20,
# ylim=c(0,1),
# cex=1.3,
# xlab='log(dose)',
# ylab='Proporção de insetos mortos')
# lines(c(-Inf, seq(0.95, 2.5, 0.37)), prediL, lwd=2, col = 3)
```

Ajustando o modelo com duas funções de ligação diferentes, concluímos que o $\log(dose)$ interfere na eficiência do inseticida. Isto é, sua presença no modelo foi estatisticamente significativa.

```
# doses letais de 50% e 90%
m5$fitted.values
```

```
##          1          2          3          4          5
## 4.683272e-09 1.277012e-01 3.222009e-01 5.422763e-01 8.097071e-01
##          6
## 9.105789e-01
```

```
log10(da5$dose)
```

```
## [1] -2.0000000 0.4149733 0.5797836 0.7075702 0.8864907 1.0086002
```

```
# Tem-se, por exemplo, que a chance de morte dos insetos sob a
# dose de inseticida 5.1 ( $\log_{10}(dose) = 0.7075702$ )
# foi  $\exp(7.1462 * (0.7075702 - 0.5797836))$  ??? 2,5 vezes a daqueles sob a
# dose 3.8 ( $\log_{10}(dose) = 0.5797836$ ).
```

```
# (c) LD50 ??? 4.8
10^(-m5$coef[1]/m5$coef[2])
```

```
## (Intercept)
## 4.828918
```

```
10^((log(0.9/0.1)+4.8869)/7.1462)
```

```
## [1] 9.801973
```

```
9.801973
```

```
## [1] 9.801973
```



```
# ??????????????????????/
# 0.5 = exp(-4.890 + 3.105 x)/(1 + exp(-4.890 + 3.105 x))
# 0.5 (1 + exp(-4.890 + 3.105 x)) = exp(-4.890 + 3.105 x)
# # log: log(0.5) + 0.5 (-4.890 + 3.105 x) = -4.890 + 3.105 x
# log(0.5) # -0.6931472
# -4.89/2
# 7.1462/2
# # -0.6931472 - 2.445 + 3.5731 = (3.105-1.5525)x
# -0.6931472 - 2.445 + 3.5731
# 3.105-1.5525
# # x = 1.5525/1.751853
# x = 1.5525/0.4349528
# x
# dl50 <- exp(x)
# dl50
# *****
```

6. Indivíduos hipertensos participaram de um estudo que teve por objetivo avaliar o efeito de dietas e medicamentos na redução da pressão arterial diastólica (PAD). Os dados estão na Tabela 5.

```
# i <- expand.grid(Dieta = c("Usual", "-Gordura", "-Sal"),
#                 Medicamento = c("Placebo", "Chlortalidone",
#                                   "Atenolol"),
#                 Sim = 0, Não = 0, Total = 0)
# da6 <- edit(i)
# str(da6)
# write.table(da6, file = "da6.txt")
da6 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/ex6dietas.txt",
                  header = TRUE)
names(da6) <- c("Dieta", "Medicamento", "Sim", "Não")
print(xtable(da6, type = "html", caption = ""))
```

% latex table generated in R 3.3.1 by xtable 1.8-2 package % Wed Nov 16 17:47:51 2016

	Dieta	Medicamento	Sim	Não
1	usual	placebo	23	67
2	usual	chlor	47	40
3	usual	atenolol	61	26
4	rg	placebo	40	50
5	rg	chlor	65	22
6	rg	atenolol	64	24
7	rs	placebo	33	45
8	rs	chlor	57	32
9	rs	atenolol	61	29

Tabela 3:

- Represente os dados graficamente.
- Analise os dados e apresente conclusões

```

tab1 <- xtabs(Sim~Medicamento + Dieta, data = da6)
tab2 <- xtabs(Não~Medicamento + Dieta, data = da6)

# gráfico de mosaico para os dados
library(mosaic)
par(mfrow=c(1, 2), cex = 0.8)

mosaicplot(tab1, col = c("skyblue", "salmon", "orange"),
            main = "Sim")
mosaicplot(tab2, col = c("skyblue", "salmon", "orange"),
            main = "Não")

```

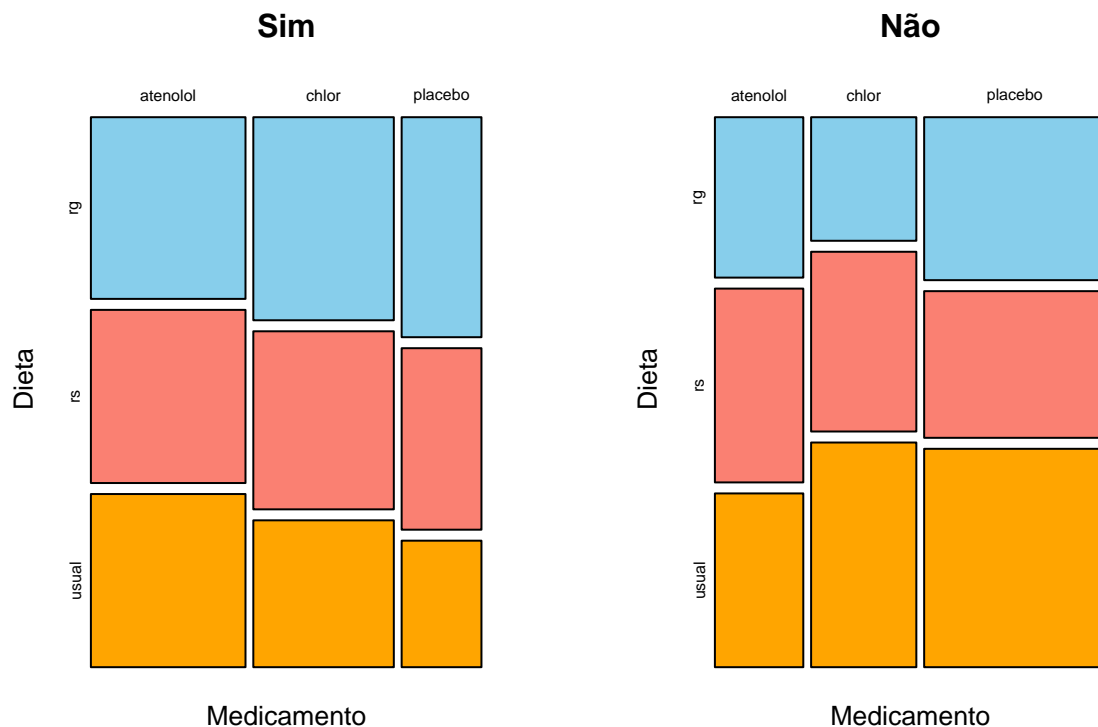


Figura 1: Diminuição na Pressão Arterial

```

m6 <- glm(cbind(Sim, Não) ~ Dieta + Medicamento,
          data = da6,
          family = binomial(link = "logit"))
summary(m6)

```

```

##
## Call:
## glm(formula = cbind(Sim, Não) ~ Dieta + Medicamento, family = binomial(link = "logit"),
##      data = da6)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## -0.88060 -0.49160  1.36334  0.01514  0.75138 -0.79134  0.85668
##      8      9

```

```
## -0.19022 -0.65129
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.1731    0.1779   6.593 4.30e-11 ***
## Dietars          -0.2813    0.1885  -1.492 0.135641
## Dietausual       -0.6335    0.1860  -3.405 0.000661 ***
## Medicamentochlor -0.2724    0.1874  -1.454 0.145958
## Medicamentoplacebo -1.3995    0.1880  -7.445 9.73e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 83.1284  on 8  degrees of freedom
## Residual deviance:  5.2611  on 4  degrees of freedom
## AIC: 58.386
##
## Number of Fisher Scoring iterations: 3
```

```
anova(m6, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(Sim, Não)
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                8      83.128
## Dieta                2    11.125      6     72.004  0.00384 **
## Medicamento        2    66.743      4      5.261 3.214e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(car)
```

```
##
## Attaching package: 'car'

## The following objects are masked from 'package:mosaic':
##
##   deltaMethod, logit

## The following object is masked from 'package:dplyr':
##
##   recode
```

```
# Anova de inserção de variáveis não sequencial
Anova(m6)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: cbind(Sim, Não)
##           LR Chisq Df Pr(>Chisq)
## Dieta      11.827  2  0.002702 **
## Medicamento 66.743  2  3.214e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notamos acima que, como já era possível de imaginar através dos gráficos em mosaico, as variáveis foram significativas. A *Anova* realizada reenforça essa hipótese, dado que a inserção não sequencial de variáveis também mostrou que elas são relevantes. As interações foram pouco ou nada relevantes para o modelo.

Odds: A chance de um indivíduo com restrição de gordura ter redução na pressão em relação à dieta de restrição de sódio é:

```
1/exp(m6$coefficients[2])
```

```
## Dietars
## 1.324829
```

```
# Embora as dietas rs e rg não tenham apresentado diferenças estatísticas
# significativas, foi observado indícios de que a chance de redução da
# PAD seja um um pouco maior sob a dieta com restrição de gordura do que
# sob a dieta com restrição de sal.
```

```
# RG e usual
1/exp(m6$coefficients[3])
```

```
## Dietausual
## 1.884206
```

Lista VI

1. Os dados na Tabela são de um estudo sobre doença respiratória crônica em que as categorias da variável resposta Y indicam: I = sem sintomas, II = tosse por menos de 3 meses ao ano, III = tosse por mais de 3 meses ao ano e IV = tosse e outros sintomas por mais de 3 meses ao ano. O objetivo do estudo foi investigar a associação da poluição do ar, poluição no trabalho e status de fumo com a doença.

```
# da7 <- read.table("da7.txt")
da7 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/respiratoria.txt",
                  header = TRUE)
print(xtable(da7, type = "html", caption = ""))
```

% latex table generated in R 3.3.1 by xtable 1.8-2 package % Wed Nov 16 17:47:56 2016

	n1	n2	n3	n4	ar	trab	fumo
1	158	9	4	1	B	N	N
2	167	19	5	3	B	N	E
3	307	102	83	68	B	N	S
4	26	5	5	1	B	S	N
5	38	12	4	4	B	S	E
6	94	48	46	60	B	S	S
7	94	7	5	1	A	N	N
8	67	8	4	3	A	N	E
9	184	65	33	36	A	N	S
10	32	3	6	1	A	S	N
11	39	11	4	2	A	S	E
12	77	48	39	51	A	S	S

Tabela 4:

- Analise os dados por meio do modelo **logitos cumulativos**.
- Avalie a qualidade de adequação do modelo considerado em (a).
- Apresente conclusões sobre a associação de interesse

```
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
##
```

```
## Attaching package: 'VGAM'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##      logit
```

```
## The following objects are masked from 'package:mosaic':
```

```
##
```

```
##      chisq, logit
```

```
names(da7) <- c("I", "II", "III", "IV", "Ar", "Trabalho", "Fumo")
```

```
# não houve efeito significativo da variável de poluição no ar
```

```
# ajuste de um modelo logitos cumulativos com chances proporcionais
```

```
m7.1 <- vglm(cbind(I, II, III, IV) ~ Fumo + Trabalho,
             data = da7, family = cumulative(parallel = TRUE,
                                              reverse = FALSE))
```

```
coef(m7.1, matrix = TRUE)
```

```
##              logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)      1.7008674      2.5821548      3.5007771
## FumoN            0.4013478      0.4013478      0.4013478
## FumoS           -1.4505706     -1.4505706     -1.4505706
## TrabalhoS       -0.8591378     -0.8591378     -0.8591378
```

```
# ajuste de um modelo logitos cumulativos sem chances proporcionais
m7.2 <- vglm(cbind(I, II, III, IV) ~ Fumo + Trabalho,
             data = da7, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))
coef(m7.2, matrix = TRUE)
```

```
##           logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)      1.6612419      2.84694620      3.8067147
## FumoN           0.4316733      0.03705781      0.9731269
## FumoS          -1.3995398     -1.72342341     -1.7783494
## TrabalhoS       -0.8416610     -0.85802821     -0.9043969
```

```
TRV <- 2*(logLik(m7.2)-logLik(m7.1))
gl <- (length(coef(m7.2))-length(coef(m7.1)))
p <- 1-pchisq(TRV, gl)
cbind(TRV, gl, p)
```

```
##           TRV gl           p
## [1,] 12.3538  6 0.05452564
```

```
# p [mais ou menos] significativo
```

```
# deve-se descobrir qual das variáveis esta interferindo
# e precisa de chances proporcionais (parcial)
```

```
m7.1 <- vglm(cbind(I, II, III, IV) ~ Fumo,
             data = da7, family = cumulative(parallel = TRUE,
                                             reverse = FALSE))
```

```
# ajuste de um modelo logitos cumulativos sem chances proporcionais
```

```
m7.2 <- vglm(cbind(I, II, III, IV) ~ Fumo,
             data = da7, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))
```

```
TRV <- 2*(logLik(m7.2)-logLik(m7.1))
gl <- (length(coef(m7.2))-length(coef(m7.1)))
p <- 1-pchisq(TRV, gl)
cbind(TRV, gl, p) # significativo - o fumo precisa de chances
```

```
##           TRV gl           p
## [1,] 12.70856  4 0.01279123
```

```
# proporcionais
```

```
m7.1 <- vglm(cbind(I, II, III, IV) ~ Trabalho,
             data = da7, family = cumulative(parallel = TRUE,
                                             reverse = FALSE))
```

```
# ajuste de um modelo logitos cumulativos sem chances proporcionais
```

```
m7.2 <- vglm(cbind(I, II, III, IV) ~ Trabalho,
             data = da7, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))
```

```

TRV <- 2*(logLik(m7.2)-logLik(m7.1))
gl <- (length(coef(m7.2))-length(coef(m7.1)))
p <- 1-pchisq(TRV, gl)
cbind(TRV, gl, p) # não significativo

##          TRV gl          p
## [1,] 0.2551593  2 0.8802233

m7 <- vglm(cbind(I, II, III, IV) ~ Fumo + Trabalho,
           data = da7, family = cumulative(parallel = FALSE~Fumo,
                                           reverse = FALSE))

# modelos com chances proporcionais parciais para o fumo
coef(m7, matrix = TRUE)

##          logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)      1.6661554      2.84414186      3.7810307
## FumoN            0.4306765      0.03814593      0.9705752
## FumoS           -1.4014119     -1.72401437     -1.7776227
## TrabalhoS       -0.8548397     -0.85483967     -0.8548397

# verificando a adequação do modelo
rp <- resid(m7, type = "pearson")
Qp <- sum(rp^2)
cbind(Qp, 1-pchisq(Qp, 26)) # quais gl são esses?

##          Qp
## [1,] 13.67476 0.9770414

QL <- deviance(m7)
cbind(QL, 1-pchisq(QL, 26))

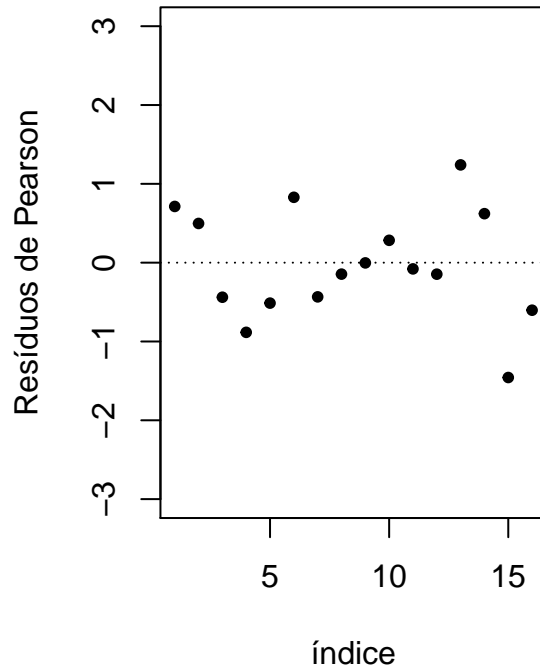
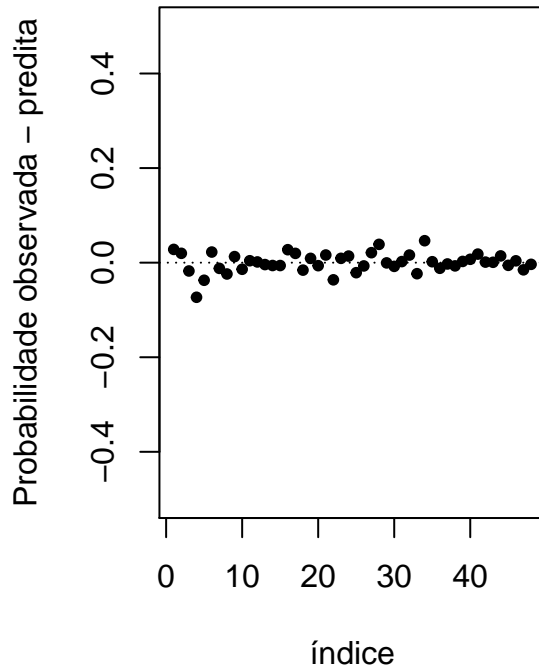
##          QL
## [1,] 13.75799 0.9760529

# os resultados são muito próximos
# gráficos de todos os resíduos dos logitos
par(mfrow=c(1,2))
rb <- m7@y - fitted(m7)
plot(1:12, rb[,1], pch=20,
     ylim=c(-0.5,0.5),xlim=c(1,48),
     xlab="índice",
     ylab="Probabilidade observada - predita")
points(13:24, rb[,2],pch=20)
points(25:36, rb[,3],pch=20)
points(37:48, rb[,4],pch=20)
abline(h=0,lty=3)

# resíduos de pearson
rp <- resid(m7, type = "pearson")

```

```
plot(1:12, rp[,1], pch=20,
     ylim=c(-3,3), xlim=c(1,16),
     xlab="índice",
     ylab="Resíduos de Pearson")
points(13:24, rp[,2], pch=20)
points(25:36, rp[,3], pch=20)
abline(h=0, lty=3)
```



os gráficos mostram um bom ajuste do modelo

Avaliamos dois modelos: o com chances proporcionais e sem. Para este problema, o modelo sem chances proporcionais parciais se mostrou necessário, dando lugar ao outro ajuste. Esta conclusão foi obtida utilizando o Teste de Razão de Verossimilhança entre os dois modelos. O modelo final é o de chances proporcionais parciais para a variável fumo.

Os resíduos de Pearson estando no intervalo entre -1 e 1 sugerem uma boa qualidade de ajuste.

A interpretação fica como abaixo: logito 1: gravidade doença respiratória I / (II ou III ou IV)

- Poluição no Trabalho - variável sem chances proporcionais $OR\ S|N = \exp(-0.855) = 0.425$
 $OR\ N|S = 1/\exp(-0.855) = 2.351$
- Ex-fumante = nível de referência (logito 1) $OR\ N | EX = \exp(0.43) = 1.537$

Chance de nenhum sintoma de doença respiratória crônica entre os não fumantes é ??? 1,5 vezes a dos ex-fumantes.

2. Os dados de um estudo sobre demência realizado com indivíduos de 65 anos ou mais de idade são mostrados na Tabela **37*. Considerando que o objetivo do estudo é o de investigar a associação entre as variáveis X_1 (uso de tabaco) e X_2 (problema cardíaco) com o estado geral de saúde dos indivíduos (variável resposta):


```
# i <- expand.grid(Tabaco = c("Sim", "Nao"),
#                 Cardiac = c("Nao", "Sim"),
#                 Exc = 0, Bom = 0, Mod = 0, Ruim = 0,
#                 Total = 0)
# da8 <- edit(i)
# str(da8)
# write.table(da8, file = "da8.txt")

da8 <- read.table("https://docs.ufpr.br/~giolo/CE073/Dados/idosos.txt",
                  header = TRUE)
print(xtable(da8, type = "html", caption = ""))
```

% latex table generated in R 3.3.1 by xtable 1.8-2 package % Wed Nov 16 17:48:00 2016

	tabaco	pcard	exc	bom	mod	ruim
1	Sim	Sim	27	76	101	39
2	Sim	Nao	402	1050	522	145
3	Nao	Sim	83	406	442	114
4	Nao	Nao	1959	4521	2243	405

Tabela 5:

- Represente graficamente os dados do estudo.
- Análise os dados fazendo uso do modelo logitos cumulativos.
- Apresente conclusões sobre a associação de interesse.

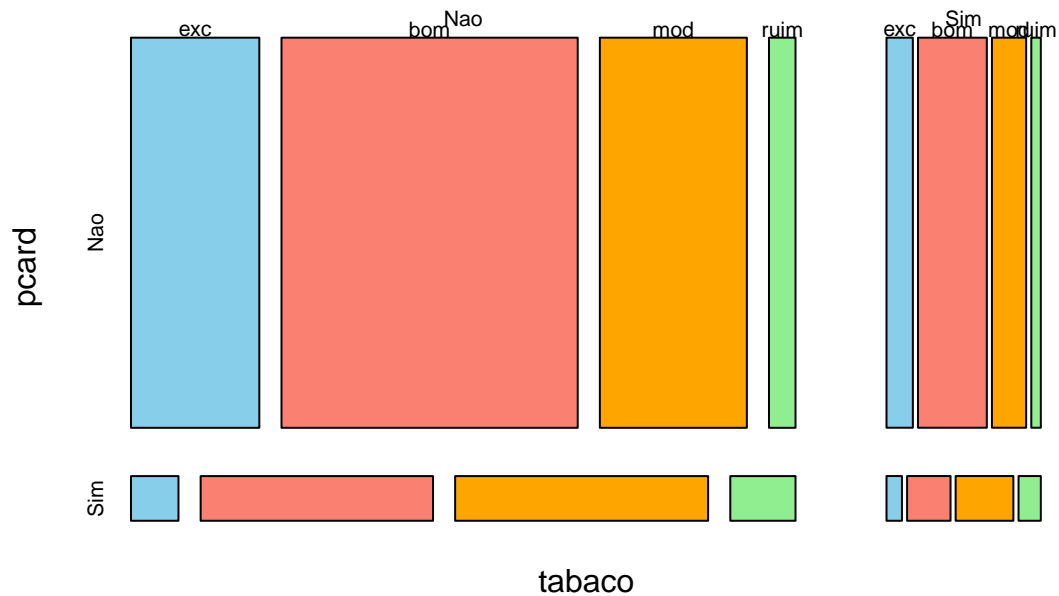
```
str(da8)
```

```
## 'data.frame': 4 obs. of 6 variables:
## $ tabaco: Factor w/ 2 levels "Nao","Sim": 2 2 1 1
## $ pcard : Factor w/ 2 levels "Nao","Sim": 2 1 2 1
## $ exc : int 27 402 83 1959
## $ bom : int 76 1050 406 4521
## $ mod : int 101 522 442 2243
## $ ruim : int 39 145 114 405
```

```
tab <- xtabs(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
            data = da8)

# gráfico de mosaico para os dados
mosaicplot(tab,
            col = c("skyblue", "salmon", "orange", "lightgreen"),
            main = "Estado geral de saúde")
```

Estado geral de saúde



```
# ajuste dos modelos
# logitos cumulativos com chances proporcionais
m8.1 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = cumulative(parallel = TRUE,
                                             reverse = FALSE))
coef(m8.1, matrix = TRUE)
```

```
##               logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)   -1.2984333      0.9012630      3.0116663
## tabacoSim      -0.1536361     -0.1536361     -0.1536361
## pcardSim       -1.0259096     -1.0259096     -1.0259096
```

```
# logitos cumulativos sem chances proporcionais
m8.2 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))
coef(m8.2, matrix = TRUE)
```

```
##               logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)   -1.3020308      0.8973008      3.0690885
## tabacoSim      -0.1274031     -0.1286946     -0.4579608
## pcardSim       -1.0467222     -1.0358216     -0.9646038
```

```
# teste da razão das verossimilhanças
TRV <- 2*(logLik(m8.2)-logLik(m8.1))
gl <- length(coef(m8.2))-length(coef(m8.1))
p <- 1 - pchisq(TRV, gl)
cbind(TRV, gl, p) # p significativo
```

```
##               TRV gl      p
## [1,] 15.11377  4 0.004470974
```

```

# deve-se descobrir qual das variáveis esta interferindo
# e precisa de chances proporcionais (parcial)

m8.1 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = cumulative(parallel = TRUE,
                                             reverse = FALSE))

# ajuste de um modelo logitos cumulativos sem chances proporcionais
m8.2 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))

TRV <- 2*(logLik(m8.2)-logLik(m8.1))
gl <- (length(coef(m8.2))-length(coef(m8.1)))
p <- 1-pchisq(TRV, gl)
cbind(TRV, gl, p) # significativo - o tabaco precisa de chances

```

```

##          TRV gl          p
## [1,] 14.88761  2 0.0005850543

```

```

# proporcionais

#-----
m8.1 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = cumulative(parallel = TRUE,
                                             reverse = FALSE))

# ajuste de um modelo logitos cumulativos sem chances proporcionais
m8.2 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = cumulative(parallel = FALSE,
                                             reverse = FALSE))

TRV <- 2*(logLik(m8.2)-logLik(m8.1))
gl <- (length(coef(m8.2))-length(coef(m8.1)))
p <- 1-pchisq(TRV, gl)
cbind(TRV, gl, p) # não significativo

```

```

##          TRV gl          p
## [1,] 0.5879971  2 0.7452776

```

```

m8 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
           data = da8, family = cumulative(parallel = FALSE~tabaco,
                                           reverse = FALSE))

```

```

# modelos com chances proporcionais parciais para o TABACO
coef(m8, matrix = TRUE)

```

```

##          logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)      -1.3030321      0.8967713      3.0826522
## tabacoSim        -0.1271912     -0.1283847     -0.4581315
## pcardSim         -1.0253390     -1.0253390     -1.0253390

```

```
rp <-resid(m8, type = "pearson")
Qp <-sum(rp^2)
cbind(Qp, 1-pchisq(Qp, 5)) # quais gl são esses?
```

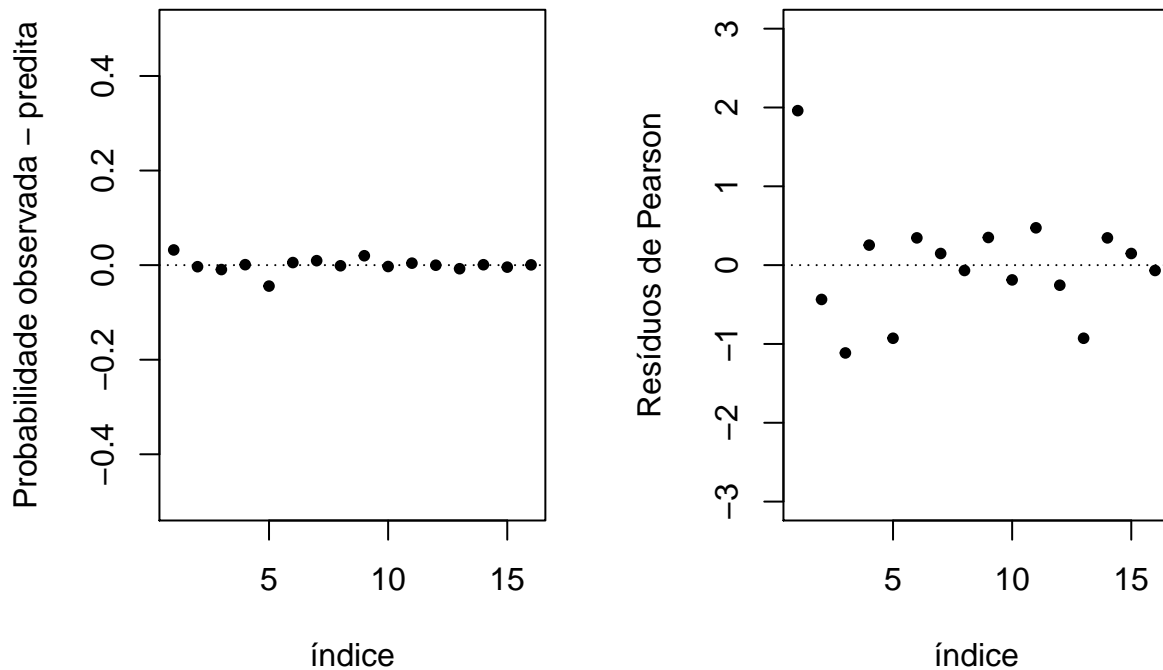
```
##                Qp
## [1,] 6.791648 0.2366026
```

```
QL <- deviance(m8)
cbind(QL, 1-pchisq(QL, 5))
```

```
##                QL
## [1,] 6.528623 0.2581221
```

```
# os resultados são muito próximos
# gráficos de todos os resíduos dos logitos
par(mfrow=c(1,2))
rb <- m8@y - fitted(m8)
plot(1:4, rb[,1], pch=20,
     ylim=c(-0.5,0.5), xlim=c(1, 16),
     xlab="índice",
     ylab="Probabilidade observada - predita")
points(5:8, rb[,2],pch=20)
points(9:12, rb[,3],pch=20)
points(13:16, rb[,4],pch=20)
abline(h=0,lty=3)

# resíduos de pearson
rp <-resid(m8, type = "pearson")
plot(1:4, rp[,1], pch=20,
     ylim=c(-3,3), xlim=c(1,16),
     xlab="índice",
     ylab="Resíduos de Pearson")
points(5:8, rp[,2],pch=20)
points(9:12, rp[,3],pch=20)
points(13:16, rp[,4],pch=20)
abline(h=0,lty=3)
```



```
coef(m8, matrix = TRUE)
```

```
##          logit(P[Y<=1]) logit(P[Y<=2]) logit(P[Y<=3])
## (Intercept)    -1.3030321      0.8967713      3.0826522
## tabacoSim      -0.1271912     -0.1283847     -0.4581315
## pcardSim       -1.0253390     -1.0253390     -1.0253390
```

```
# os gráficos mostram um bom ajuste do modelo
```

Avaliamos dois modelos: o com chances proporcionais e sem. Para este problema, o modelo sem chances proporcionais parciais se mostrou necessário, dando lugar ao outro ajuste. Esta conclusão foi obtida utilizando o Teste de Razão de Verossimilhança entre os modelos. O modelo final é o de chances proporcionais parciais para a variável tabaco.

Os resíduos de Pearson estando no intervalo entre **-2** e **2** sugerem uma boa qualidade de ajuste.

A interpretação fica como abaixo: logito 1: estado de saúde Exc / (Bom ou Mod ou Ruim)

- Tabaco - variável com chances proporcionais $OR\ S|N = \exp(-0.1271912) = 0.8805653$
 $OR\ N|S = 1/\exp(-0.855) = 1.135634$

Interpretação: A chance de alguém que não fuma estar na classe excelente de saúde é apenas um pouco maior do que a das pessoas que consomem tabaco.

- Ausência de problema cardíaco = nível de referência (logito 1) $OR\ S|N = \exp(-1.0253390) = 0.3586749$
 $OR\ N|S = 1/\exp(-1.0253390) = 2.78804$

Interpretação: A chance de alguém que não tem problemas cardíacos estar na classe excelente de saúde é 2.79 vezes a chance de quem tem problemas cardíacos.

3. Analise os dados do estudo dispostos na Tabela 37 por meio:

- (a) do modelo logitos categorias adjacentes;
- (b) do modelo logitos razão contínua.

```
# modelo logitos categorias adjacentes
```

```
# com chances proporcionais
```

```
m8.3 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = acat(parallel = TRUE,
                                       reverse = TRUE))
coef(m8.3, matrix = TRUE)
```

```
##           loge(P[Y=1]/P[Y=2]) loge(P[Y=2]/P[Y=3]) loge(P[Y=3]/P[Y=4])
## (Intercept)      -0.8347761          0.7010264          1.7030790
## tabacoSim        -0.1135151         -0.1135151         -0.1135151
## pcardSim         -0.6587568         -0.6587568         -0.6587568
```

```
# sem chances proporcionais
```

```
m8.4 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = acat(parallel = FALSE,
                                       reverse = TRUE))
coef(m8.4, matrix = TRUE)
```

```
##           loge(P[Y=1]/P[Y=2]) loge(P[Y=2]/P[Y=3]) loge(P[Y=3]/P[Y=4])
## (Intercept)      -0.84303147          0.70411995          1.7108833
## tabacoSim        -0.08578843         -0.01907128         -0.4264225
## pcardSim         -0.61955710         -0.81983308         -0.3516252
```

```
TRV <- deviance(m8.3) - deviance(m8.4)
gl <- df.residual(m8.3) - df.residual(m8.4)
p <- 1 - pchisq(TRV, gl)
cbind(TRV, gl, p) # p significativo - um modelo com chances
```

```
##           TRV gl           p
## [1,] 22.58513  4 0.0001532303
```

```
# proporcionais é necessário
```

```
#----- só tabaco
```

```
# com chances proporcionais
```

```
m8.3 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = acat(parallel = TRUE,
                                       reverse = TRUE))
```

```
# sem chances proporcionais
```

```
m8.4 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = acat(parallel = FALSE,
                                       reverse = TRUE))
```

```
TRV <- deviance(m8.3) - deviance(m8.4)
gl <- df.residual(m8.3) - df.residual(m8.4)
p <- 1 - pchisq(TRV, gl)
cbind(TRV, gl, p) # p significativo
```

```
##          TRV gl          p
## [1,] 11.08578  2 0.003915192
```

```
#----- só pcard
# com chances proporcionais
m8.3 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = acat(parallel = TRUE,
                                       reverse = TRUE))

# sem chances proporcionais
m8.4 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = acat(parallel = FALSE,
                                       reverse = TRUE))

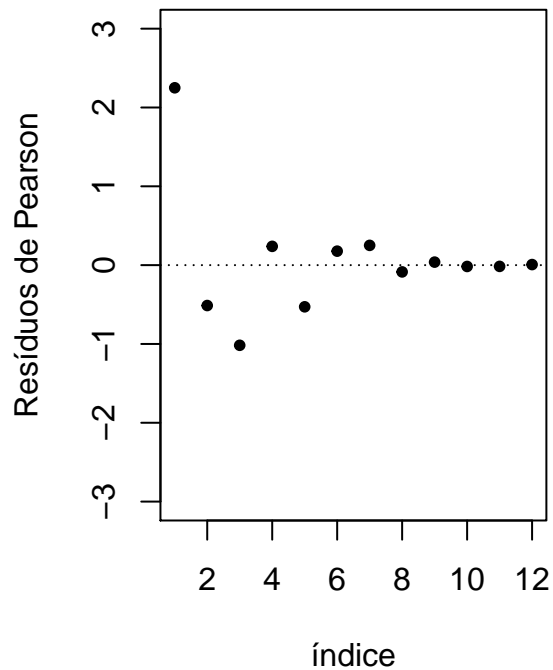
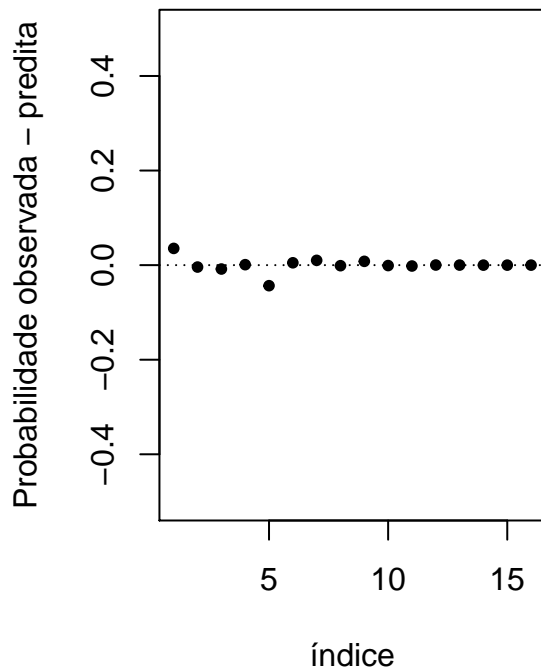
TRV <- deviance(m8.3) - deviance(m8.4)
gl <- df.residual(m8.3) - df.residual(m8.4)
p <- 1 - pchisq(TRV, gl)
cbind(TRV, gl, p) # p significativo
```

```
##          TRV gl          p
## [1,] 11.53136  2 0.003133257
```

```
# as chances propocionais são necessárias para ambas as variáveis
# modelo final
m8.4 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = acat(parallel = FALSE,
                                       reverse = TRUE))

#----- avaliação do modelo
par(mfrow=c(1,2))
rb <- m8.4@y - fitted(m8.4)
plot(1:4, rb[,1], pch=20,
     ylim = c(-0.5,0.5),
     xlim = c(1,16),
     xlab = "índice",
     ylab = "Probabilidade observada - predita")
points(5:8, rb[,2], pch=20)
points(9:12, rb[,3], pch=20)
points(13:16, rb[,4], pch=20)
abline(h=0, lty=3)

rp <- resid(m8.4, type = "pearson")
plot(1:4, rp[,1],
     pch = 20,
     ylim = c(-3,3),
     xlim = c(1,12),
     xlab = "índice", ylab="Resíduos de Pearson")
points(5:8, rp[,2], pch=20)
points(9:12, rp[,3], pch=20)
abline(h=0, lty=3)
```



```
coef(m8.4, matrix = TRUE)
```

```
##           loge(P[Y=1]/P[Y=2]) loge(P[Y=2]/P[Y=3]) loge(P[Y=3]/P[Y=4])
## (Intercept)      -0.84303147      0.70411995      1.7108833
## tabacoSim        -0.08578843     -0.01907128     -0.4264225
## pcardSim         -0.61955710     -0.81983308     -0.3516252
```

Para esse modelo tem-se a partir do logito 3, por exemplo, que a chance de estado de saúde moderado (em relação ao estado de saúde ruim) dentre os indivíduos idosos sem problema cardíaco foi $1/\exp(-0.3516252) = 1.421376$ vezes a daqueles com problema cardíaco. Por outro lado, a chance de estado de saúde moderado (em relação ao estado de saúde ruim) dentre os indivíduos idosos que não utilizavam tabaco foi $1/\exp(-0.4264225) = 1.5317681$ vezes a daqueles que utilizavam.

```
# modelo logitos razão contínua
# sem chances proporcionais
m8.5 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = cratio(parallel = FALSE,
                                         reverse = FALSE))

# com chances proporcionais
m8.6 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = cratio(parallel = TRUE,
                                         reverse = FALSE))

coef(m8.6, matrix = TRUE)
```

```
##           logit(P[Y>1|Y>=1]) logit(P[Y>2|Y>=2]) logit(P[Y>3|Y>=3])
## (Intercept)      1.3107074      -0.5394842      -1.7516802
## tabacoSim        0.1554012      0.1554012      0.1554012
## pcardSim         0.8053110      0.8053110      0.8053110
```



```
TRV <- 2*(logLik(m8.5)-logLik(m8.6))
gl <- length(coef(m8.5))-length(coef(m8.6))
p <- 1-pchisq(TRV,gl)
cbind(TRV, gl, p) # há necessidade de chances proporcionais
```

```
##          TRV gl          p
## [1,] 37.07079  4 1.741816e-07
```

```
# tabaco
# SEM chances proporcionais
m8.5 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = cratio(parallel = FALSE,
                                         reverse = FALSE))
```

```
# chances proporcionais
m8.6 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco,
             data = da8, family = cratio(parallel = TRUE,
                                         reverse = FALSE))
```

```
TRV <- 2*(logLik(m8.5)-logLik(m8.6))
gl <- length(coef(m8.5))-length(coef(m8.6))
p <- 1-pchisq(TRV,gl)
cbind(TRV, gl, p) # não há necessidade de chances proporcionais
```

```
##          TRV gl          p
## [1,] 9.08337  2 0.01065544
```

```
# pcard
# SEM chances proporcionais
m8.5 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = cratio(parallel = FALSE,
                                         reverse = FALSE))
```

```
# chances proporcionais
m8.6 <- vglm(cbind(exc, bom, mod, ruim) ~ pcard,
             data = da8, family = cratio(parallel = TRUE,
                                         reverse = FALSE))
```

```
TRV <- 2*(logLik(m8.5)-logLik(m8.6))
gl <- length(coef(m8.5))-length(coef(m8.6))
p <- 1-pchisq(TRV,gl)
cbind(TRV, gl, p) # não há necessidade de chances proporcionais
```

```
##          TRV gl          p
## [1,] 28.21272  2 7.476286e-07
```

```
# modelo final
m8.5 <- vglm(cbind(exc, bom, mod, ruim) ~ tabaco + pcard,
             data = da8, family = cratio(parallel = FALSE,
                                         reverse = FALSE))
```

```
coef(m8.5, matrix = TRUE)
```

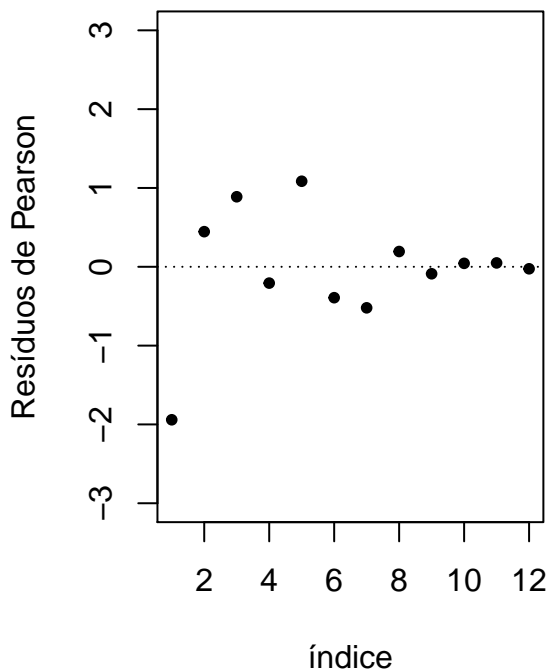
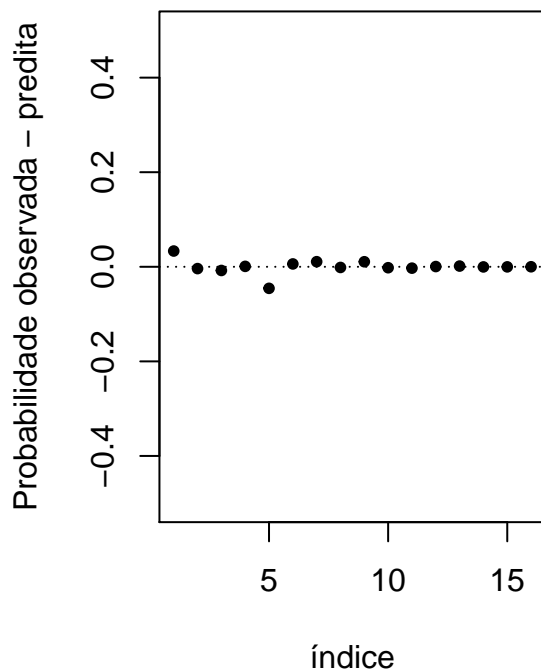
```
##          logit(P[Y>1|Y>=1]) logit(P[Y>2|Y>=2]) logit(P[Y>3|Y>=3])
## (Intercept)          1.302607          -0.5396741          -1.7103210
## tabacoSim            0.124586            0.1052933            0.4252645
## pcardSim            1.045950            0.8879319            0.3500455
```

```
QL <- deviance(m8.5)
cbind(QL, 1-pchisq(QL, 3))
```

```
##          QL
## [1,] 6.124047 0.1057286
```

```
par(mfrow=c(1,2))
rb <- m8.5@y - fitted(m8.5)
plot(1:4,rb[,1], pch=20,
     ylim = c(-0.5,0.5),
     xlim = c(1,16),
     xlab = "índice",
     ylab = "Probabilidade observada - predita")
points(5:8, rb[,2], pch=20)
points(9:12, rb[,3], pch=20)
points(13:16, rb[,4], pch=20)
abline(h=0,lty=3)
```

```
rp <- resid(m8.5, type = "pearson")
plot(1:4, rp[,1], pch=20,
     ylim=c(-3,3),
     xlim=c(1,12),
     xlab="índice",
     ylab="Resíduos de Pearson")
points(5:8, rp[,2],pch=20)
points(9:12, rp[,3],pch=20)
abline(h=0,lty=3)
```



```
str(da8)
```

```
## 'data.frame':  4 obs. of  6 variables:
## $ tabaco: Factor w/ 2 levels "Nao","Sim": 2 2 1 1
## $ pcard : Factor w/ 2 levels "Nao","Sim": 2 1 2 1
## $ exc   : int  27 402 83 1959
## $ bom   : int  76 1050 406 4521
## $ mod   : int  101 522 442 2243
## $ ruim  : int  39 145 114 405
```

A partir do logito 1 tem-se, por exemplo, que a chance de estado de saúde excelente (em relação aos demais estados) dentre os indivíduos idosos sem problema cardíaco foi $\exp(1.045950) \approx 2.846101$ vezes a daqueles com problema cardíaco. Por outro lado, a chance de estado de saúde excelente (em relação aos demais estados) dentre os indivíduos idosos que não utilizavam tabaco foi $\exp(0.124586) = 1.132679$ vezes a daqueles que utilizavam.

4. (...)

Lista VII

1. Um estudo caso-controle com pareamento 1:1 envolveu um total de 78 pessoas a fim de pesquisar se fumo (SMK) estaria associado ao infarto do miocárdio (MI) (Kleinbaum, 1994). As variáveis consideradas no pareamento foram: idade, raça e sexo. Duas outras variáveis não consideradas no pareamento foram: pressão arterial sistólica (SBP) e status do eletrocardiograma (ECG).

Estudo caso-controle com pareamento 1:1 (pares informativos: aqueles em que $x_{i1} \neq x_{i2}$)

- (a) Ajuste um modelo de regressão logística condicional aos dados.
- (b) Apresente conclusões com base no modelo ajustado.

```
library(survival)

dados <- read.table("infart.txt",
                    header = TRUE)

model <- clogit(MI ~ (SMK + SBP + ECG)^2 + strata(par),
                data = dados)

model
```

```
## Call:
## clogit(MI ~ (SMK + SBP + ECG)^2 + strata(par), data = dados)
##
##              coef exp(coef) se(coef)      z      p
## SMK        -6.81713  0.00109  7.41058 -0.92 0.36
## SBP         0.02045  1.02066  0.02740  0.75 0.46
## ECG         3.75271 42.63644  7.88380  0.48 0.63
## SMK:SBP     0.05828  1.06001  0.05633  1.03 0.30
## SMK:ECG    -1.47174  0.22953  1.96530 -0.75 0.45
```

```
## SBP:ECG -0.00834  0.99169  0.05118 -0.16 0.87
##
## Likelihood ratio test=12.7 on 6 df, p=0.0487
## n= 78, number of events= 39
```

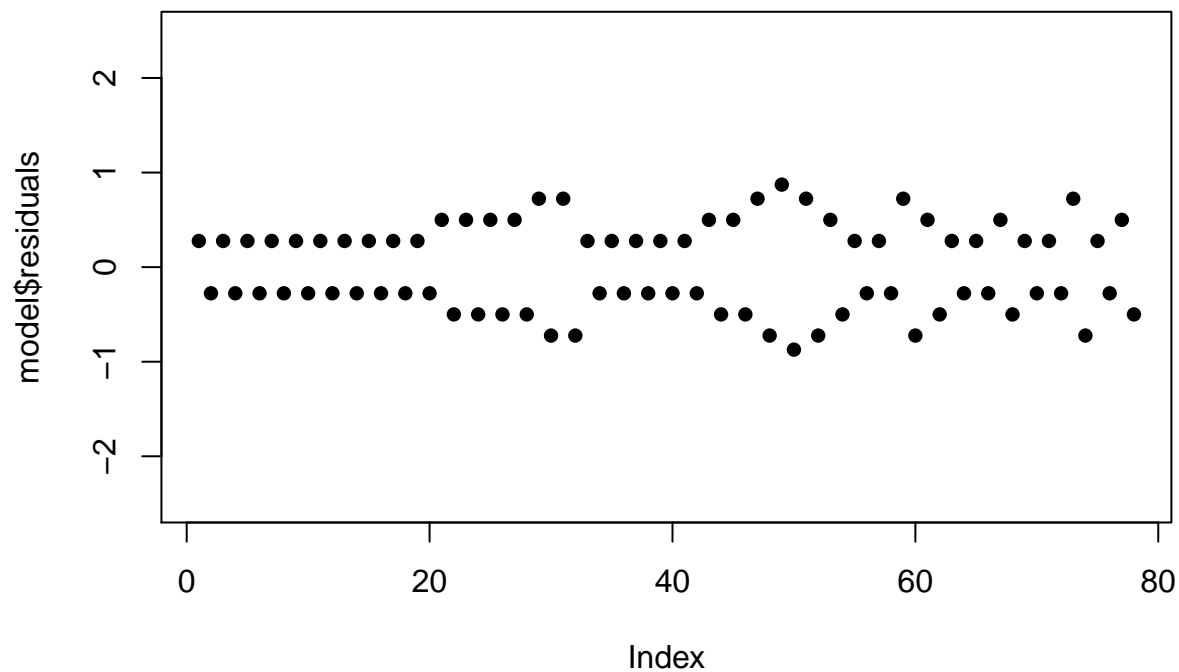
```
model <- clogit (MI ~ SMK + SBP + ECG + strata(par),
                data = dados)
model
```

```
## Call:
## clogit(MI ~ SMK + SBP + ECG + strata(par), data = dados)
##
##      coef exp(coef) se(coef)      z      p
## SMK 0.837      2.309      0.652 1.28 0.199
## SBP 0.035      1.036      0.021 1.66 0.096
## ECG 1.831      6.242      1.154 1.59 0.113
##
## Likelihood ratio test=11.4 on 3 df, p=0.00963
## n= 78, number of events= 39
```

```
model <- clogit (MI ~ SBP + strata(par),
                data = dados)
summary(model)
```

```
## Call:
## coxph(formula = Surv(rep(1, 78L), MI) ~ SBP + strata(par), data = dados,
##       method = "exact")
##
##      n= 78, number of events= 39
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## SBP 0.04807  1.04924  0.02033 2.364  0.0181 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## SBP      1.049      0.9531      1.008      1.092
##
## Rsquare= 0.079 (max possible= 0.5 )
## Likelihood ratio test= 6.41 on 1 df,  p=0.01132
## Wald test              = 5.59 on 1 df,  p=0.01807
## Score (logrank) test = 6.13 on 1 df,  p=0.01333
```

```
plot(model$residuals, pch = 16, ylim = c(-2.5,2.5))
```



Para os dados analisados foram encontradas evidências de que SBP está associada com infarto do miocárdio (MI), uma vez que a chance de se observar pacientes com SBP elevado foi maior entre os casos (com MI) do que entre os controles (sem MI).

2. Para investigar se a infertilidade secundária feminina estaria associada com abortos prévios (induzidos ou espontâneos), 83 mulheres com a infertilidade mencionada e 83 sem a infertilidade participaram de um estudo caso-controle em que foram pareadas por: idade, número de gravidez (incluindo abortos) e anos de escolaridade (Trichopoulos et al., 1976).