Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019)

Exercises of Chapter 6: Applying Logistic Regression

Kimmo Vehkalahti and Brian S. Everitt

10 November 2018

Exercises

Exercise 6.1

Use the womensrole data (see below) and modify the R code given in the **Examples of Chapter 6** to analyse and visualize the data as instructed.

```
womensrole <- structure(list(
    education = c(0L, 1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L, 11L, 12L, 13L, 14L, 15L,
                  16L, 17L, 18L, 19L, 20L, 0L, 1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L,
                  11L, 12L, 13L, 14L, 15L, 16L, 17L, 18L, 19L, 20L),
  sex = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
                    1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
                    2L, 2L, 2L, 2L, 2L, 2L, 2L),
                  .Label = c("Male", "Female"), class = "factor"),
  agree = c(4, 2, 4, 6, 5, 13, 25, 27, 75, 29, 32, 36, 115, 31, 28, 9, 15, 3, 1, 2, 3, 4,
            1, 0, 6, 10, 14, 17, 26, 91, 30, 55, 50, 190, 17, 18, 7, 13, 3, 0, 1, 2),
  disagree = c(2, 0, 0, 3, 5, 7, 9, 15, 49, 29, 45, 59, 245, 70, 79, 23, 110, 29, 28, 13, 20,
               2, 0, 0, 1, 0, 7, 5, 16, 36, 35, 67, 62, 403, 92, 81, 34, 115, 28, 21, 2, 4)),
.Names = c("education", "sex", "agree", "disagree"),

row.names = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12", "13", "14", "15",
"16", "17", "18", "19", "20", "21", "22", "23", "24", "25", "26", "27", "28", "29", "30", "31",
"32", "33", "34", "35", "36", "37", "38", "39", "40", "41", "42"), class = "data.frame")

head(womensrole)
```

```
##   education  sex agree disagree
## 1         0 Male     4        2
## 2         1 Male     2        0
## 3         2 Male     4        0
## 4         3 Male     6        3
## 5         4 Male     5        5
## 6         5 Male    13        7
```

Exercise 6.3

Use the accidents data (see below) and modify the R code given in the **Examples of Chapter 6** to analyse the data as instructed.

```
accidents <- structure(list(

  weight = structure(c(1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L),
                     .Label = c("Small", "Standard"), class = "factor"),

  eject = structure(c(1L, 1L, 2L, 2L, 1L, 1L, 2L, 2L),
                    .Label = c("No", "Yes"), class = "factor"),

  type = structure(c(1L, 2L, 1L, 2L, 1L, 2L, 1L, 2L),
                   .Label = c("Collision", "Rollover"), class = "factor"),

  severely = c(150, 112, 23, 80, 1022, 404, 161, 265),

  nseverely = c(350, 60, 26, 19, 1878, 148, 111, 22)),

.Names = c("weight", "eject", "type", "severely", "nseverely"),
row.names = c(NA, -8L), class = "data.frame")

head(accidents)
```

```
##     weight eject      type severely nseverely
## 1    Small    No Collision      150       350
## 2    Small    No  Rollover      112        60
## 3    Small   Yes Collision       23        26
## 4    Small   Yes  Rollover       80        19
## 5 Standard    No Collision     1022      1878
## 6 Standard    No  Rollover      404       148
```

```
tail(accidents)
```

```
##     weight eject      type severely nseverely
## 3    Small   Yes Collision       23        26
## 4    Small   Yes  Rollover       80        19
## 5 Standard    No Collision     1022      1878
## 6 Standard    No  Rollover      404       148
## 7 Standard   Yes Collision      161       111
## 8 Standard   Yes  Rollover      265        22
```

Exercise 6.4

Use the statistics data (see below) and modify the R code given in the **Examples of Chapter 6** to analyse the data as instructed.

```
statistics <- structure(list(

  result = structure(c(1L, 1L, 2L, 1L, 2L, 2L, 2L, 2L, 2L, 2L,
                       2L, 2L, 2L, 2L, 2L, 1L, 2L, 1L, 1L, 2L),
                     .Label = c("Fail", "Pass"), class = "factor"),

  test = c(525, 533, 545, 582, 581, 576, 572, 609, 559, 543,
           576, 525, 574, 582, 574, 471, 595, 557, 557, 584),

  grade = structure(c(2L, 3L, 2L, 1L, 3L, 4L, 2L, 1L, 3L, 4L,
                      2L, 1L, 4L, 3L, 2L, 2L, 3L, 1L, 1L, 1L),
                    .Label = c("A", "B", "C", "D"), class = "factor")),

.Names = c("result", "test", "grade"), row.names = c(NA, -20L), class = "data.frame")

head(statistics)
```

```
##   result test grade
## 1   Fail  525     B
## 2   Fail  533     C
## 3   Pass  545     B
## 4   Fail  582     A
## 5   Pass  581     C
## 6   Pass  576     D
```

```
tail(statistics)
```

```
##    result test grade
## 15   Pass  574     B
## 16   Fail  471     B
## 17   Pass  595     C
## 18   Fail  557     A
## 19   Fail  557     A
## 20   Pass  584     A
```

Exercise 6.5

Use the menstruation data (see below) and modify the R code given in the **Examples of Chapter 6** to analyse and visualize the data as instructed.

```
menstruation <- structure(list(
  age = c(11.08, 11.33, 11.58, 11.83, 12.08, 12.33, 12.58, 12.83, 13.08,
          13.33, 13.58, 14.08, 14.33, 14.58, 15.08, 15.33, 15.58, 17.58),

  bmens = c(2, 5, 10, 17, 16, 29, 39, 51, 47, 67, 81, 79, 90, 93, 117, 107, 92, 1049),

  n = c(120, 88, 105, 111, 100, 93, 100, 108, 99, 106, 117,
        98, 97, 100, 122, 111, 94, 1049)),

.Names = c("age", "bmens", "n"), row.names = c(NA, -18L), class = "data.frame")

menstruation
```

```
##       age bmens    n
## 1   11.08     2  120
## 2   11.33     5   88
## 3   11.58    10  105
## 4   11.83    17  111
## 5   12.08    16  100
## 6   12.33    29   93
## 7   12.58    39  100
## 8   12.83    51  108
## 9   13.08    47   99
## 10  13.33    67  106
## 11  13.58    81  117
## 12  14.08    79   98
## 13  14.33    90   97
## 14  14.58    93  100
## 15  15.08   117  122
## 16  15.33   107  111
## 17  15.58    92   94
## 18  17.58  1049 1049
```

Exercise 6.6

Use the gambling data (see below) and modify the R code given in the **Examples of Chapter 6** to analyse and visualize the data as instructed. Below you will find some information on the variables of the data. For more details, see the original article:

Donati, M. A., Chiesi, F. and Primi, C. (2013). A model to explain at-risk/problem gambling among male and female adolescents: Gender similarities and differences. *Journal of Adolescence*, 36, 129-137.

```
gambling <- read.table("data/gambling.txt", sep = '\t', header = TRUE)

head(gambling)
```

```
##   Id Gender      Age G1 G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 GamblF GFalla
## 1  1      2 13.41667  0  1  0  0  0  0  1  0  0   0   0      2      5
## 2  2      1 13.83333  0  0  0  0  1  0  0  0  2   2   0      5      2
## 3  3      2 13.83333  1  0  0  0  0  0  0  0  3   1   0      5      3
## 4  4      2 13.83333  0  0  0  0  0  0  0  0  1   0   0      1      6
## 5  5      2 13.83333  1  0  0  0  0  3  0  0  0   0   0      4      6
## 6  6      2 13.83333  3  1  2  0  0  1  1  2  4   0   0     14      2
##   SensaS SuperS EconG PareG PeerG Nsymp P3prob P2prob
## 1     21     18    30     0     0     0      0      0
## 2     36     11    35     1    NA     0      0      0
## 3     17     11    39     1     0     0      0      0
## 4     24     26    38     0     0     0      0      0
## 5     25     24    36     1     0     1      1      1
## 6     34     18    26     1     1     3      2      1
```

- Gender: 1 = male, 2 = female
- Age: Age in years
- G1: Cards (0 = never, 1 = less than monthly, 2 = monthly, 3 = weekly, 4 = daily)
- G2: Coin toss
- G3: Bets on games of skill
- G4: Bets on sport games
- G5: Bets on horse races
- G6: Bingo
- G7: Dice games
- G8: Slot machines
- G9: Instant scratch-cards
- G10: Lotteries
- G11: Online games
- GamblF: Gambling frequency (sum of G1 to G11)
- GFalla: Probabilistic reasoning (Gambler's fallacy)
- SensaS: Sensation seeking
- SuperS: Superstitious thinking
- EconG: Economic perception of gambling (high score = more negative attitude)
- PareG: Parental gambling behavior
- PeerG: Peer gambling behavior
- Nsymp: Problem gambling (total of symptoms)
- P3prob: Problem gambling (0 = non-problem, 1 = at-risk, 2 = problem)
- P2prob: Problem gambling (0 = non-problem, 1 = at-risk/problem)