Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019)
Exercises of Chapter 7: Survival Analysis

Kimmo Vehkalahti and Brian S. Everitt

15 November 2018

Exercises

Exercise 7.2

Use the cancer data (see below) and modify the R code given in the **Examples of Chapter 7** to analyse and visualize the data as instructed.

```
cancer <- structure(list(
  time = c(23, 47, 69, 70, 100, 101, 148, 181, 198, 208, 212, 224, 5, 8, 10, 13, 18,
           24, 26, 26, 31, 35, 40, 41, 48, 50, 59, 61, 68, 71, 76, 105, 107, 109,
           113, 116, 118, 143, 145, 162, 188, 212, 217, 225),
  event = c(1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
            1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0),
  metastized = structure(c(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, 2L, 2L, 2L, 2L,
                           2L, 2L, 2L, 2L, 2L),
                 .Label = c("No", "Yes"), class = "factor")),
  .Names = c("time", "event", "metastized"),
row.names = c(NA, -44L), class = "data.frame")

str(cancer)
```

```
## 'data.frame':    44 obs. of  3 variables:
##  $ time      : num  23 47 69 70 100 101 148 181 198 208 ...
##  $ event     : num  1 1 1 0 0 0 1 1 0 0 ...
##  $ metastized: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(cancer)
```

```
##   time event metastized
## 1   23     1         No
## 2   47     1         No
## 3   69     1         No
## 4   70     0         No
## 5  100     0         No
## 6  101     0         No
```

Exercise 7.3

Use the glioma data (see below) and modify the R code given in the **Examples of Chapter 7** to analyse the data as instructed.

```
glioma <- structure(list(
  no. = c(1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L, 11L, 12L, 13L, 14L, 15L, 16L,
          17L, 18L, 19L, 1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L, 11L, 12L, 13L,
          14L, 15L, 16L, 17L, 18L),
  age = c(41L, 45L, 48L, 54L, 40L, 31L, 53L, 49L, 36L, 52L, 57L, 55L, 70L, 39L,
          40L, 47L, 58L, 40L, 36L, 27L, 32L, 53L, 46L, 33L, 19L, 32L, 70L, 72L,
          46L, 44L, 83L, 57L, 71L, 61L, 65L, 50L, 42L),
  sex = structure(c(1L, 1L, 2L, 2L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 1L, 2L, 1L, 1L,
                    1L, 2L, 1L, 2L, 2L, 2L, 1L, 2L, 1L, 1L, 1L, 2L, 2L, 2L, 2L,
                    1L, 1L, 1L, 2L, 2L, 2L, 1L),
          .Label = c("Female", "Male"), class = "factor"),
  histology = structure(c(2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 1L, 1L,
                          1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 1L,
                          1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L),
          .Label = c("GBM", "Grade3"), class = "factor"),
  group = structure(c(2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
                      2L, 2L, 2L, 2L, 2L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
                      1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L),
          .Label = c("Control", "RIT"), class = "factor"),
  event = c(TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE,
            FALSE, FALSE, TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE,
            TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
            TRUE, TRUE, TRUE, TRUE, TRUE),
  time = c(53L, 28L, 69L, 58L, 54L, 25L, 51L, 61L, 57L, 57L, 50L, 43L, 20L, 14L,
           36L, 59L, 31L, 14L, 36L, 34L, 32L, 9L, 19L, 50L, 48L, 8L, 8L, 11L, 12L,
           15L, 5L, 8L, 8L, 6L, 14L, 13L, 25L)),

  .Names = c("no.", "age", "sex", "histology", "group", "event", "time"),
  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"), class = "data.frame")

str(glioma)
```

```
## 'data.frame':    37 obs. of  7 variables:
##  $ no.      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ age      : int  41 45 48 54 40 31 53 49 36 52 ...
##  $ sex      : Factor w/ 2 levels "Female","Male": 1 1 2 2 1 2 2 2 2 2 ...
##  $ histology: Factor w/ 2 levels "GBM","Grade3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ group    : Factor w/ 2 levels "Control","RIT": 2 2 2 2 2 2 2 2 2 2 ...
##  $ event    : logi  TRUE FALSE FALSE FALSE FALSE TRUE ...
##  $ time     : int  53 28 69 58 54 25 51 61 57 57 ...
```

Exercise 7.4

Use the teeth data (see below) and modify the R code given in the **Examples of Chapter 7** to visualize and analyse the data as instructed. Below you will find some information on the variables of the data. For more details, see the original article:

Palotie, U., Eronen, A. K., Vehkalahti, K. and Vehkalahti, M. M. (2017). Longevity of 2- and 3-surface restorations in posterior teeth of 25- to 30-year-olds attending Public Dental Service—A 13-year observation. *Journal of Dentistry*, 62, 13-17.

The teeth of these data have been coded using ISO dental notation system, see:

https://en.wikipedia.org/wiki/Dental_notation#ISO_System_by_the_World_Health_Organization

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

```
##   ID Gender Age ISOcode Reint Survyear Type4 Material Type2
## 1  1      1  30      24     0    2.510     1        2     1
## 2  2      1  30      26     0    2.510     3        2     2
## 3  3      1  30      16     1    0.301     3        2     2
## 4  4      2  30      14     1    0.545     1        2     1
## 5  5      2  30      15     1    9.033     1        2     1
## 6  6      1  28      15     1   12.526     1        2     1
```

```
tail(teeth)
```

```
##        ID Gender Age ISOcode Reint Survyear Type4 Material Type2
## 5537 5537      2  25      14     1    0.348     1        2     1
## 5538 5538      2  25      36     0    0.504     4        2     2
## 5539 5539      2  25      37     0    0.504     4        2     2
## 5540 5540      1  25      27     1    9.732     3        2     2
## 5541 5541      1  25      46     1    9.438     4        2     2
## 5542 5542      2  25      36     0    9.068     4        2     2
```

```
# Set maximum longevity to 13 years and censor all at the end of 13th year:
teeth$Survyear[teeth$Survyear > 13] <- 13
teeth$Reint[teeth$Survyear == 13] <- 0
```

- ID: Tooth ID
- Gender: 1=Male, 2=Female
- Age: Patient age in years (2002)
- ISOcode: Tooth code (ISO)
- Reint: Re-intervention: 0=No, 1=Yes
- Survyear: Longevity of restoration in years
- Type4: 1=Upper premolar, 2=Lower premolar, 3=Upper molar, 4=Lower molar
- Material: 1=Amalgam, 2=Composite
- Type2: 1=Premolar, 2=Molar