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

Kimmo Vehkalahti and Brian S. Everitt

15 November 2018

Solutions

Exercise 7.2

```
brcancer <- 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, 2L, 2L),
                .Label = c("No", "Yes"), class = "factor")),
  .Names = c("time", "event", "metastized"),
row.names = c(NA, -44L), class = "data.frame")

str(brcancer)
```

```
## '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(brcancer)
```

```
##   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
```

```
attach(brcancer)
library(survival)

#plot of survival function
plot(survfit(Surv(time, event) ~ metastized), xlab = "Time", ylab = "Proportion", lty = 1:2)
legend("bottomleft", c("No","Yes"), lty=1:2)
```
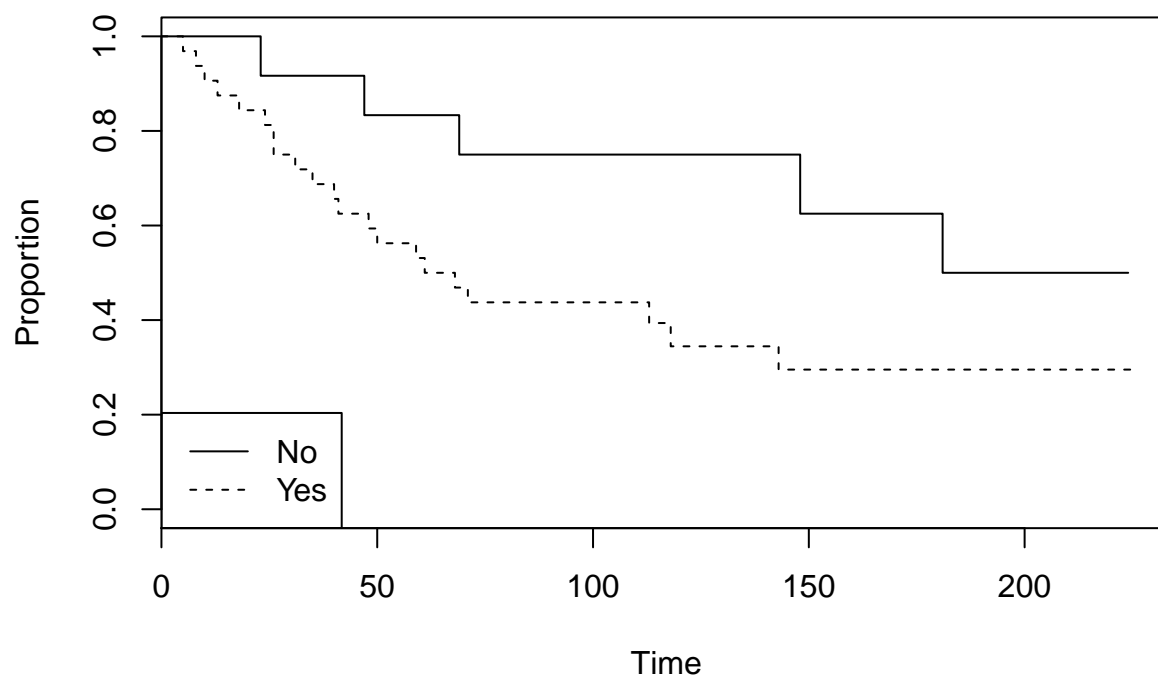
```
#logrank test
survdiff(Surv(time,event) ~ metastized)
```

```
## Call:
## survdiff(formula = Surv(time, event) ~ metastized)
##
##                 N Observed Expected (O-E)^2/E (O-E)^2/V
## metastized=No  12        5      9.2      1.91      3.04
## metastized=Yes 32       21     16.8      1.05      3.04
##
##  Chisq= 3  on 1 degrees of freedom, p= 0.0813
```

```
detach(brcancer)
```

Exercise 7.3

```r
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 ...
```

```r
attach(glioma)

library(survival)
glioma_cox <- coxph(Surv(time, event) ~ age + sex + histology + group)
summary(glioma_cox)
```

```
## Call:
## coxph(formula = Surv(time, event) ~ age + sex + histology + group)
```

```
## 
##   n= 37, number of events= 23
## 
##                    coef exp(coef) se(coef)       z Pr(>|z|)
## age             0.04344   1.04440  0.02152   2.019   0.0435 *
## sexMale         0.06411   1.06621  0.49755   0.129   0.8975
## histologyGrade3 -1.48480   0.22655  0.61826  -2.402   0.0163 *
## groupRIT        -2.44441   0.08678  0.58156  -4.203 2.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
##                 exp(coef) exp(-coef) lower .95 upper .95
## age               1.04440     0.9575   1.00127    1.0894
## sexMale           1.06621     0.9379   0.40210    2.8272
## histologyGrade3   0.22655     4.4141   0.06744    0.7611
## groupRIT          0.08678    11.5238   0.02776    0.2713
## 
## Concordance= 0.855  (se = 0.068 )
## Rsquare= 0.636   (max possible= 0.981 )
## Likelihood ratio test= 37.39  on 4 df,   p=1.497e-07
## Wald test            = 27.32  on 4 df,   p=1.714e-05
## Score (logrank) test = 40.28  on 4 df,   p=3.788e-08
detach(glioma)
```

Required CI given in summary above - hazard on radioimmunotherapy (RIT) is between about 3% and 27% of hazard on standard therapy.

You should plot the survival curves of RIT and standard treatments.

Exercise 7.4

```r
teeth <- read.csv(file = "data/teeth.csv", row.names = 1)
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
```

```r
#tail(teeth)

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

As an example, Figure 2 of the original article (Palotie et al., 2017), entitled, "Survival (%) of restorations placed in posterior teeth (n = 5542) of 25- to 30-year-olds illustrated as medians and Kaplan-Meier curves with 95% confidence intervals by type of tooth", was created by one of the authors, *Anna Eronen*, as follows:

```r
library(survival)
#install.packages("plotrix")
library(plotrix)

plot(survfit(Surv(Survyear, Reint) ~ strata(Type2), data = teeth), conf.int = "both",
     col = c("red", "blue"), lwd = 1.7, axes = FALSE, yscale = 100,
     main = "", ylab = "", xlab = "", cex.axis = 1.2)
title(xlab = "Length of follow-up in years", line = 1.7, cex.lab = 1.2)
title(ylab = "Survivals %", line = 2.5, cex.lab = 1.2)
#
ablineclip(v=2, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(v=4, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(v=6, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(v=8, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(v=10, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(v=12, y1=0, y2=1.0, col="darkgrey", lty=3, lwd=1.4)
ablineclip(h=0.2, x1=0, x2=13, col="darkgrey", lty=3, lwd=1.4)
ablineclip(h=0.4, x1=0, x2=13, col="darkgrey", lty=3, lwd=1.4)
ablineclip(h=0.6, x1=0, x2=13, col="darkgrey", lty=3, lwd=1.4)
ablineclip(h=0.8, x1=0, x2=13, col="darkgrey", lty=3, lwd=1.4)
ablineclip(h=1.0, x1=0, x2=13, col="black",lty=1,lwd=1.0)
ablineclip(v=13, y1=0,y2=1,col="black",lty=1,lwd=1.0)
axis(1, at=c(0,1,3,5,7,9,11,13), labels=c(0,1,3,5,7,9,11,13), pos=0, las=1)
axis(2, at=c(0.0,0.2,0.4,0.6,0.8,1.0), labels=c(0,20,40,60,80,100), pos=0, las=2)
legend(x=0.15, y=0.15, c("Premolar","Molar"), col=c("red","blue"), lty=c(1,1), bty="n")
#
ablineclip(h=0.5, x1=0, x2=13, col="black", lty=5)
ablineclip(v=12.0, y1=0, y2=0.5, col="red", lty=5, lwd=1.7)
ablineclip(v=8.7, y1=0, y2=0.5, col="blue", lty=5, lwd=1.7)
```
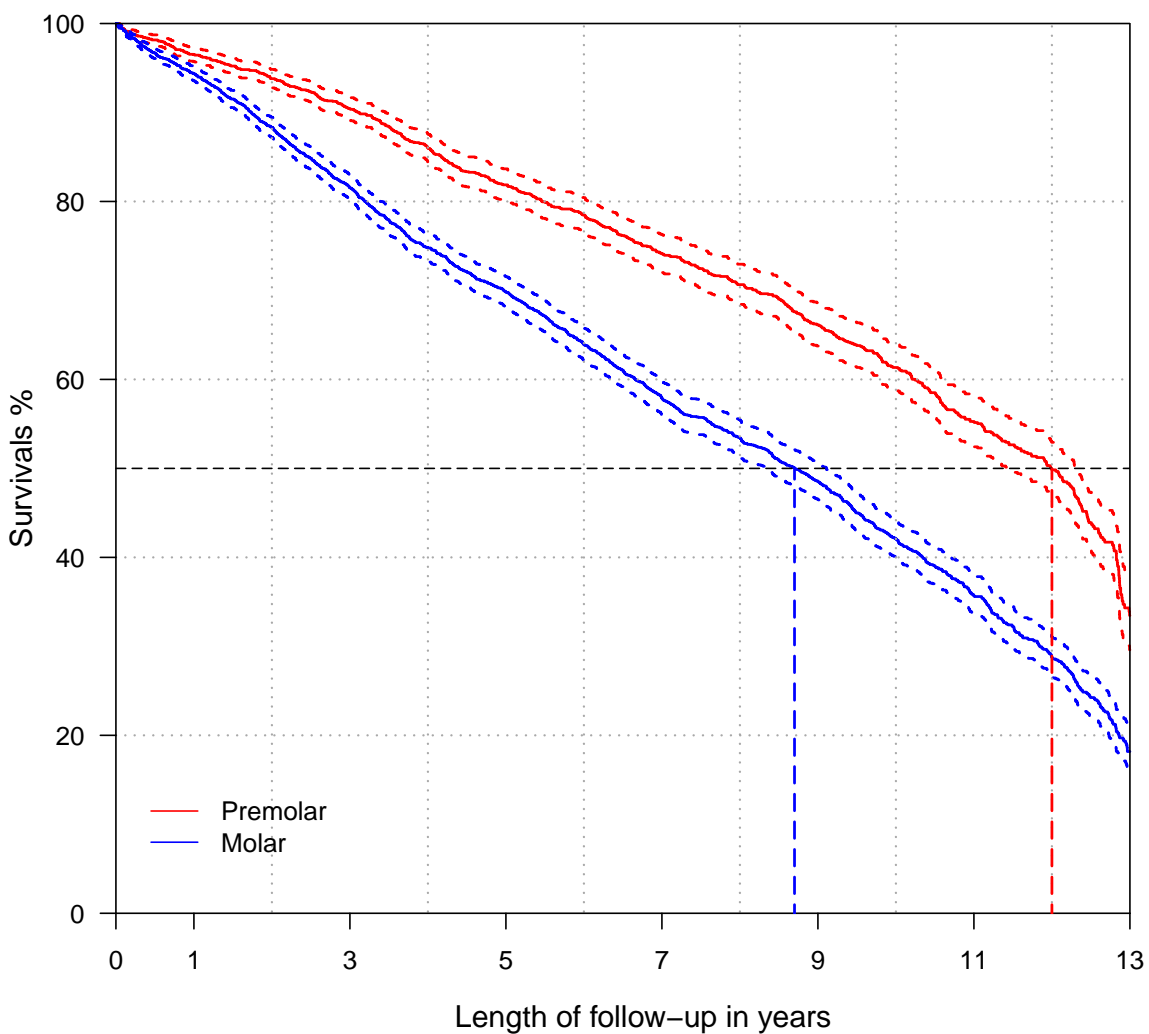
```
# example of a log-rank test:
kmtest.teeth <- survdiff(Surv(Survyear, Reint) ~ Type2, data = teeth)
kmtest.teeth

## Call:
## survdiff(formula = Surv(Survyear, Reint) ~ Type2, data = teeth)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## Type2=1 2216      746     1060      92.8       167
## Type2=2 3326     1654     1340      73.3       167
##
##  Chisq= 167  on 1 degrees of freedom, p= 0
```