Màster universitari en Estadística i Investigació Operativa

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UNIVERSITAT DE BARCELONA

Lifetime Data Analysis, Course 2020/21

Exercises Topics 3 and 4: Solutions.

Exercise 1 ($0.4 + 0.4 + 0.4 + 0.3 = 1.5$ points)

The following R output provides some information on the survival times (measured in days) in an randomized clinical trial that compared two treatments for a certain cancer:

```
> survfit(Surv(stime, status) ~ treat, cancer)

                    n events median 0.95LCL 0.95UCL
treat=Treatment 1 13      7    638     268      NA
treat=Treatment 2 13      5     NA     475      NA

> summary(survfit(Surv(stime, status) ~ treat, cancer),
+         times = seq(180, 900, 180))

                treat=Treatment 1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  180     10       3    0.769   0.117        0.571        1.000
  360      8       2    0.615   0.135        0.400        0.946
  540      5       1    0.538   0.138        0.326        0.891
  720      4       1    0.431   0.147        0.221        0.840
  900      2       0    0.431   0.147        0.221        0.840
 1080      1       0    0.431   0.147        0.221        0.840

                treat=Treatment 2
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  180     13       0    1.000  0.0000        1.000        1.000
  360     12       1    0.923  0.0739        0.789        1.000
  540      7       3    0.658  0.1407        0.433        1.000
  720      6       1    0.564  0.1488        0.336        0.946
  900      3       0    0.564  0.1488        0.336        0.946
 1080      3       0    0.564  0.1488        0.336        0.946
```

(a) Why does the `survfit` function not provide an estimation of the median survival time under Treatment 2?

(b) Why does the `survfit` function not provide an upper limit of the 95% confidence interval of the median survival time under Treatment 1?

(c) Give an interpretation of the values in the columns `n.risk`, `n.event`, and `survival` at times 180, 360, and 540 days under Treatment 1. For the ease of interpretation, you may assume that a month has 30 days.

(d) What can you deduce from the output of function `summary.survfit` on the maximum uncensored survival time under Treatment 1?

Solution:

(a) The estimator of the median survival time is the minimum over all times t such that $\hat{S}(t) \leq 0.5$: $\hat{t}_{0.5} = \min\{t : \hat{S}(t) \leq 0.5\}$. However, under Treatment 2, the estimated survival function does not fall below 0.5; for this reason, there is no (nonparametric) estimation of the median.

(b) The 95% confidence interval of the median consists of all those values of t such that

$$-1.96 \leq \frac{\hat{S}(t) - 0.5}{\sqrt{\hat{\mathrm{V}}[\hat{S}(t)]}} \leq 1.96.$$

Since $\hat{S}(t)$ is decreasing in t, the upper limit of the confidence interval corresponds to the smallest value t_s such that $\frac{\hat{S}(t_s)-0.5}{\sqrt{\hat{\mathrm{V}}[\hat{S}(t_s)]}} \leq -1.96$. Hence, if $\frac{\hat{S}(t)-0.5}{\sqrt{\hat{\mathrm{V}}[\hat{S}(t)]}} > -1.96$ for all t, the upper limit of the confidence interval is not defined.

(c) Following, we will assume that the event of interest of the clinical trial was cancer relapse. Hence, under Treatment 1, 10, 8, and 5 persons were at risk of suffering a relapse at, respectively, 6, 12, and 18 months after the start of the clinical trial. That is, these persons had neither experienced the event of interest nor were censored before. The column `n.event` indicates that during the first, second, and third 6 months of the trial, 3, 2, and 1 persons suffered a cancer relapse, respectively. The estimated probabilities to continue relapse-free after 6, 12, and 18 months are 0.769, 0.615, and 0.538, respectively.

(d) According to the output, the last cancer relapse in the Treatment 1 group was observed between 18 and 24 months after the start of the clinical trial.

Exercise 2 (0.6 + 0.9 + 1 + 1 = 3.5 points)

The file `Oral.txt` contains the survival times (in weeks) of patients with oral cancer that are classified according to their tumour type, which is either aneuploidy[1] or diploid[2].

(a) Draw the Kaplan-Meier estimates of both survival functions showing the survival time in years (assuming a year has exactly 52 weeks). What do you observe?

(b) Which are the (estimated) probabilities of surviving 1, 3, and 5 years in both study groups? Compute also the corresponding 95% confidence intervals.

(c) Compute the 95% log-transformed EP confidence bands in both study groups and present the values for 1, 3, and 5 years, respectively. Comment on the differences between the confidence bands and the confidence intervals.

(d) Choose a couple of the nonparametric tests presented on Slide 36/94 (Chapter 4) to test the hypothesis that survival is not related to the tumour type. Comment on the results and the differences between both tests.

Solution:

(a) We can observe in Figure 1 that the survival times of patients with a aneuploidy oral cancer was, on average, somewhat better.

(b) Estimation of the survival probabilities (plus computation of the 95% confidence intervals) for one, three, and five years:

```
> summary(survfit(Surv(stime, cens) ~ ttype, oralca), scale = 52,
          times = c(1, 3, 5) * 52)

Call: survfit(formula = Surv(stime, cens) ~ ttype, data = oralca)

                ttype=Aneuploidy
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1     34      18    0.654  0.0660        0.537        0.797
    3      5      11    0.381  0.0767        0.257        0.566
    5      1       2    0.229  0.0954        0.101        0.518
```

[1] https://en.wikipedia.org/wiki/Aneuploidy

[2] https://en.wikipedia.org/wiki/Ploidy\#Diploid

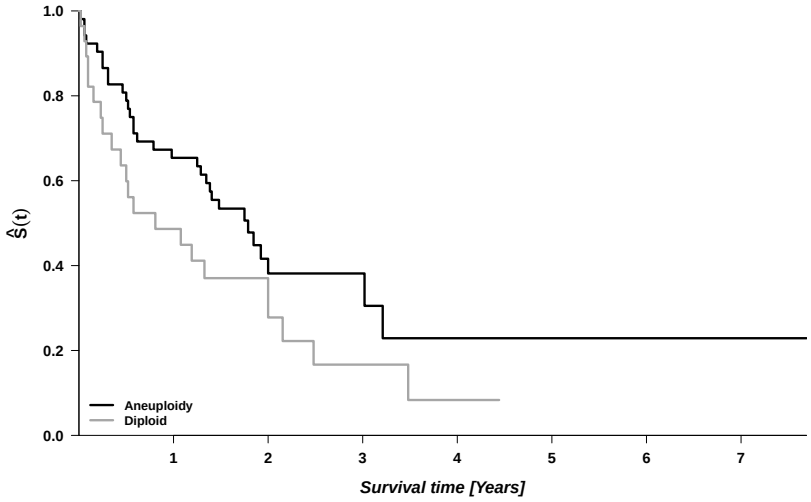

Figure 1: Estimated survival probabilities in patients with oral cancer.

```
                ttype=Diploid
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   1     13      14    0.486  0.0961       0.3302        0.716
   3      3       7    0.167  0.0815       0.0639        0.435
```

There is no estimation of the survival function for $t = 5$ (years) in the group of the diploid tumour type because the maximum survival time observed is right-censored at 231 weeks, which is less than five years.

(c) The following R code provides the 95% log-transformed EP confidence bands in both study groups after 1, 3, and 5 years. The corresponding values are shown (together with the 95% confidence intervals) in Table 1.

```
> library(km.ci)
> svforal <- survfit(Surv(stime, cens) ~ ttype, oralca)
> # Log-transformed EP confidence bands
> summary(km.ci(svforal[1], method = "logep"), scale = 52,
          times = c(1, 3, 5) * 52, extend = TRUE)
> summary(km.ci(svforal[2], method = "logep"), scale = 52,
          times = c(1, 3) * 52)
```

Table 1: 95% Confidence intervals/bands for $S(1 \text{ year})$, $S(3 \text{ y.})$, and $S(5 \text{ y.})$.

	Aneuploidy Tumour			**Diploid Tumour**		
		C. intervals/bands			**C. intervals/bands**	
Years	$\hat{\boldsymbol{S}}(\boldsymbol{t})$	KM	EP log.	$\hat{\boldsymbol{S}}(\boldsymbol{t})$	KM	EP log.
1	0.65	0.54−0.8	0.41−0.82	0.49	0.33−0.72	0.18−0.74
3	0.38	0.26−0.57	0.15−0.61	0.17	0.06−0.43	0.01−0.47
5	0.23	0.1 −0.52	0.03−0.55			

The confidence bands are larger than the confidence intervals because they cover $S(t)$ with probability 0.95 for **all** t. By contrast, the probability that at least one of the three confidence intervals for $S(1)$, $S(3)$, and $S(5)$ does not include the true probability is larger than 0.05.

(d) Following, both the logrank and the Peto-Peto test are applied. Whereas the former puts the same weight on all differences between the observed and expected numbers of events in one of the samples, the latter is more prone to detect differences at short-term.:

```
library(FHtest)
# Logrank test
FHtestrcc(Surv(stime, cens) ~ ttype, oralca)

Two-sample test for right-censored data

Parameters: rho=0, lambda=0
Distribution: counting process approach

Data: Surv(stime, cens) by ttype

                  N Observed Expected   O-E (O-E)^2/E (O-E)^2/V
ttype=Aneuploidy 52       31     36.6 -5.55     0.843      2.79
ttype=Diploid    28       22     16.4  5.55     1.873      2.79

Statistic Z= 1.7, p-value= 0.0949
Alternative hypothesis: survival functions not equal

# Peto-Peto test
> FHtestrcc(Surv(stime, cens) ~ ttype, oralca, rho = 1)

        Two-sample test for right-censored data

Parameters: rho=1, lambda=0
Distribution: counting process approach

Data: Surv(stime, cens) by ttype

                  N Observed Expected   O-E (O-E)^2/E (O-E)^2/V
ttype=Aneuploidy 52     20.2     24.4 -4.22     0.731       3.3
ttype=Diploid    28     15.1     10.9  4.22     1.643       3.3

Statistic Z= 1.8, p-value= 0.0694
Alternative hypothesis: survival functions not equal
```

There is a lot of evidence against the null hypothesis that survival does not depend on tumour type ($p = 0.0949$ and $p = 0.0694$, resp.). However, if a significance level of 0.05 is used, the evidence is not enough to reject the null hypothesis that survival does not depend on the tumour type.

Exercise 3 ($0.6 + 0.8 + 0.5 + 0.6 + 0.5 = 3$ points)

The data frame `capacitor` of the `survival` package contains data from a factorial experiment on the life of glass capacitors as a function of voltage and operating temperature. There were 8 capacitors at each combination of temperature and voltage. Testing at each combination was terminated after the fourth failure.

(a) Draw the (estimated) survival functions for both temperatures. What do you observe?

(b) Draw the (estimated) survival functions for both temperatures separately for each voltage in four different graphs with the same range of the abscissa. What do you observe?

(c) Test the hypothesis that failure times do not depend on temperature using the logrank test. What do you conclude?

(d) Test the hypothesis that failure times do not depend on temperature using the stratified logrank test. What do you conclude?

(e) Comment on the difference among both tests.

Solution:

(a) According to Figure 2, there are no big differences between both temperatures with respect to the time until failure of glass capacitors.

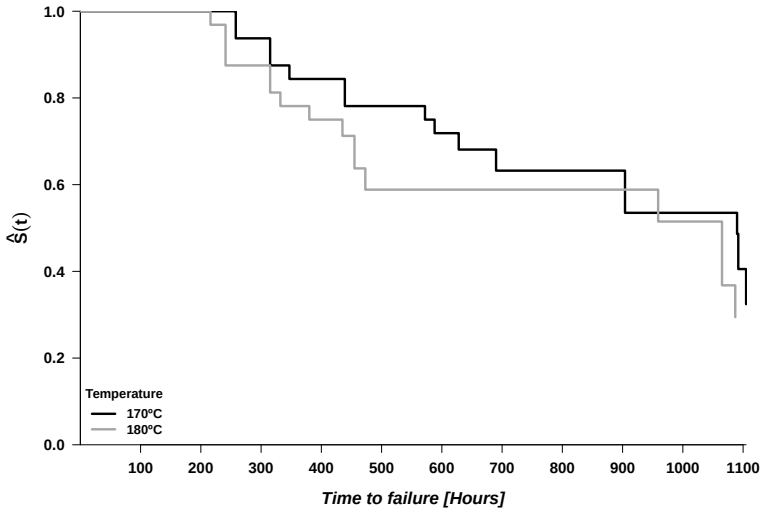

Figure 2: Survival functions of time until failure of glass capacitors.

(b) Figure 3 shows the estimated survival functions of the time until failure of glass capacitors under four different voltages. There is no clear pattern among the four voltages, even though it seems as if times until failure are, on average, somewhat larger at 170 °C. In addition, we observe that none of the survival functions falls below 0.5 since testing at each combination was terminated after the fourth failure.

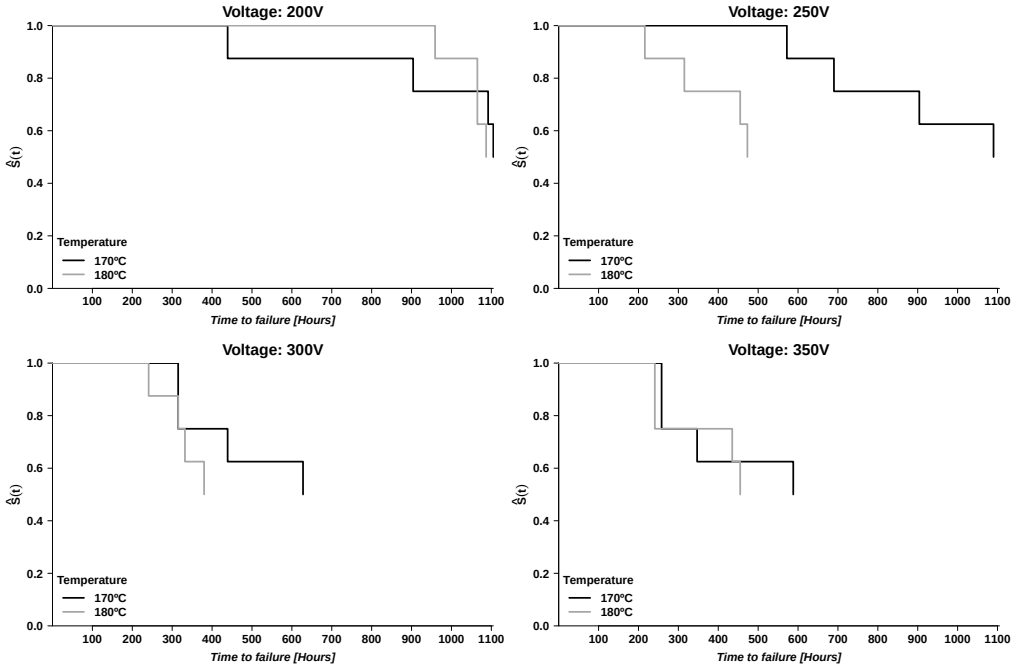

Figure 3: Survival functions of time until failure of glass capacitors under different voltages.

(c) Function `survdiff` is used to perform the logrank test comparing times to failure under both temperatures:

```
> survdiff(Surv(time, status) ~ temperature, capacitor)

Call:
survdiff(formula = Surv(time, status) ~ temperature, data = capacitor)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
temperature=170ºC 32       16     19.8     0.744      2.19
temperature=180ºC 32       16     12.2     1.214      2.19

 Chisq= 2.2  on 1 degrees of freedom, p= 0.1
```

According to the p-value obtained ($p = 0.1$), we cannot reject the null hypothesis of equal survival functions as long as a significance level of less than 0.1 is used.

(d) Function `survdiff` is used to perform the stratified logrank test comparing times to failure under both temperatures:

```
> survdiff(Surv(time, status) ~ temperature + strata(voltage), capacitor)

Call:
survdiff(formula = Surv(time, status) ~ temperature + strata(voltage),
    data = capacitor)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
temperature=170ºC 32       16     20.8      1.12      4.25
temperature=180ºC 32       16     11.2      2.09      4.25

 Chisq= 4.2  on 1 degrees of freedom, p= 0.04
```

According to the p-value ($p = 0.04$), the null hypothesis of equal survival functions can be rejected if a significance level of 0.05 is used. Given this result and based on the R output and the survival functions in Figure 3, it can be claimed that times until failure are, on average, larger at 170 °C

(e) The test result of the stratified test should be interpreted with caution since the four plots in Figure 3 do not show a clear pattern. The logrank test is optimal if survival curves do not cross, however the survival functions do cross under 200, 300, and 350 V. Moreover, a stratified test is only appropriate if the effect of temperature on the times until failure is more or less the same under all voltages under study.

Exercise 4 ($0.7 + 1.3 = 2$ points)

In a study on breast cancer with 14 women, the patients received either the treatment CMF or a placebo. The survival time of interest was the time in remission and throughout the two years of the study duration, the following survival times were observed, where '+' indicates a right-censored survival time:

Treatment	**Months in remission**
CMF	23, 24, 16^+, 18^+, 20^+, 22^+, 24^+
Placebo	15, 18, 19, 19, 20, 22^+, 24^+

(a) Draw the Nelson-Aalen estimates of both survival functions.

(b) Use the Tarone-Ware test to test the following hypothesis without using any statistical software:

$$\mathrm{H}_0\text{: } S_1(t) = S_2(t),\ \forall t, \quad \text{vs.} \quad \mathrm{H}_1\text{: } S_1(t) > S_2(t), \text{ for any } t,$$

where 1 denotes treatment CMF and 2 denotes placebo.

Solution:

(a) Both survival curves are shown in Figure 4.

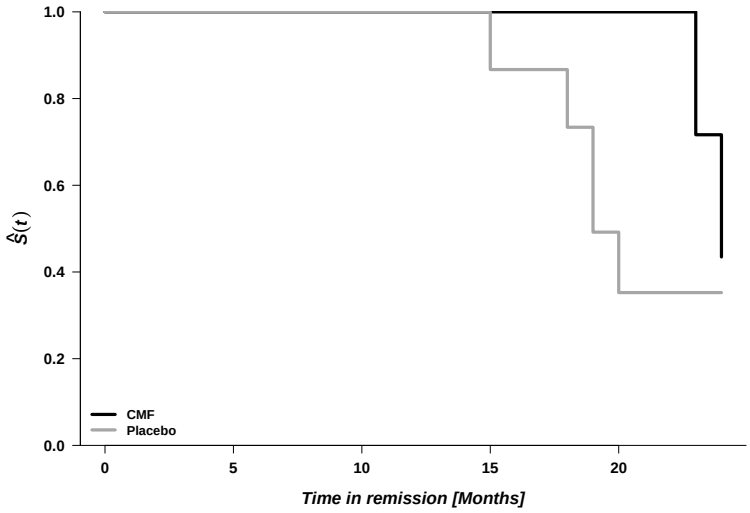

Figure 4: Survival functions of time in remission among women with breast cancer (Exercise 4).

(b) The test statistic of the Tarone-Ware test is

$$Z_W(\tau) = \frac{\sum_{i=1}^{D} \sqrt{R_i}\left(d_{i1} - R_{i1}\frac{d_i}{R_i}\right)}{\sqrt{\sum_{i=1}^{D} R_i \frac{R_{i1}}{R_i}\left(1 - \frac{R_{i1}}{R_i}\right)\frac{R_i - d_i}{R_i - 1} d_i}},$$

where $D = 6$ is the number of the different uncensored times until breast cancer relapse. Using the significance level 0.05, the null hypothesis is rejected, if $Z_W(\tau)$ is smaller than the critical value $z_{0.05} = -1.645$.

The number of events (d_{i1} and d_i) and individuals at risk (R_{i1} and R_i) of both the CMF group and the whole sample are shown in Table 2.

Table 2: Number of events and individuals at risk (Exercise 4).

Y_i	R_i	R_{i1}	d_i	d_{i1}	$d_{i1} - R_{i1}\frac{d_i}{R_i}$
15	14	7	1	0	-0.5
18	12	6	1	0	-0.5
19	10	5	2	0	-1
20	8	5	1	0	-0.625
23	4	3	1	1	0.25
24	3	2	1	1	0.33

The value of $Z_W(\tau)$:

$$\text{Numerator} = \sum_{i=1}^{D} \sqrt{R_i}\big(d_{i1} - R_{i1}\frac{d_i}{R_i}\big) = \sqrt{14}\cdot(-0.5) + \cdots + \sqrt{3}\cdot 0.33 = -7.46,$$

$$\text{Denominator} = \sqrt{\sum_{i=1}^{D} R_i \frac{R_{i1}}{R_i}\big(1 - \frac{R_{i1}}{R_i}\big)\frac{R_i - d_i}{R_i - 1}d_i} = \sqrt{14.24},$$

$$\Longrightarrow Z_W(\tau) = \frac{-7.46}{\sqrt{14.24}} = -1.98 < -1.645.$$

Hence, given $\alpha = 0.05$, the null hypothesis can be rejected. According to this result, patients with breast cancer treatment CMF have, on average, larger times in remission.

Appendix: R Code

Exercise 2

```
library(dplyr)
oralca <- read.table("OralCA.txt", skip = 7) %>%
  rename(ttype = V1, stime = V2, cens = V3) %>%
  mutate(ttype = factor(ttype, label = c("Aneuploidy", "Diploid")))

library(survival)
svforal <- survfit(Surv(stime, cens) ~ ttype, oralca)

## Figure 1
## --------
windows(width = 11)
par(font.lab = 4, font = 2, font.axis = 2, las = 1, cex.main = 1.3,
    cex.lab = 1.3, cex.axis = 1.2, mar = c(5, 5, 2, 2))
plot(svforal, xlab = "Survival time [Years]", xscale = 52, col = grey(c(0, 0.65)),
     ylab = expression(bold(hat(S)(t))), lwd = 3, cex = 1.1, yaxs = "i", xaxs = "i",
     bty = "l", xaxt = "n")
axis(1, at = 1:8 * 52, labels = 1:8)
legend("bottomleft", levels(oralca$ttype), lwd = 3, bty = "n",
       col = grey(c(0, 0.65)))
```

Exercise 3

```
capacitor$temperature <- factor(capacitor$temperature,
                                labels = paste0(c(170, 180), "ºC"))
svcap <- with(capacitor, survfit(Surv(time, status) ~ temperature))

## Figure 2
## --------
windows(width = 10, height = 7)
par(font.lab = 4, font = 2, font.axis = 2, las = 1, cex.main = 1.3,
    cex.lab = 1.3, cex.axis = 1.2, mar = c(5, 5, 2, 2))
plot(svcap, xlab = "Time to failure [Hours]", col = grey(c(0, 0.65)),
     ylab = expression(bold(hat(S)(t))), lwd = 3, cex = 1.1, yaxs = "i",
     xaxs = "i", bty = "l", xaxt = "n")
axis(1, at = seq(100, 1100, 100))
legend("bottomleft", levels(capacitor$temperature), title = "Temperature",
       lwd = 3, bty = "n", col = grey(c(0, 0.65)))
```

```
## Figure 3
## --------
xlims <- c(0, max(capacitor$time))

windows(width = 18, height = 12)
par(mfrow = c(2, 2), font.lab = 4, font = 2, font.axis = 2, las = 1,
    cex.main = 2, cex.lab = 1.5, cex.axis = 1.5, mar = c(5, 5, 3, 2))
for (i in unique(capacitor$voltage)) {
  plot(survfit(Surv(time, status) ~ temperature, subset(capacitor, voltage == i)),
       xlab = "Time to failure [Hours]", xlim = xlims, col = grey(c(0, 0.65)),
       ylab = expression(bold(hat(S)(t))), lwd = 3, cex = 1.1, yaxs = "i",
       xaxs = "i", bty = "l", xaxt = "n")
  axis(1, at = seq(100, 1100, 100))
  legend("bottomleft", levels(capacitor$temperature), title = "Temperature",
         lwd = 3, bty = "n", col = grey(c(0, 0.65)), cex = 1.5)
  title(paste0("Voltage: ", i, "V"))
}
```

Exercise 4

```
exer4 <- data.frame(treat = rep(c("Placebo", "CMF"), each = 7),
                    stime = c(15, 18, 19, 19, 20, 22, 24,
                              23, 24, 16, 18, 20, 22, 24),
                    cens = c(1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0),
                    stringsAsFactors = TRUE)

## Figure 4
## --------
windows(width = 10, height = 7)
par(font.lab = 4, font = 2, font.axis = 2, las = 1, cex.main = 1.3,
    cex.lab = 1.3, cex.axis = 1.2, mar = c(5, 5, 2, 2))
plot(survfit(Surv(stime, cens) ~ treat, exer4, stype = 2, ctype = 1),
     xlab = "Time in remission [Months]",
     ylab = expression(bolditalic(hat(S)(t))), col = grey(c(0, 0.65)), lwd = 4,
     cex = 1.1, yaxs = "i", bty = "l",
     mark.time = FALSE)
legend("bottomleft", levels(exer4$treat), lwd = 4, col = grey(c(0, 0.65)),
       bty = "n")
```