

Exercise 9 - solution

Paul Blanche

Exercise A

Question 1

We first load the data and look at the “summary”, as always.

```
rm(list=ls())
load(url("http://paulblanche.com/files/colon2.rda"))
d <- colon2
summary(d)
```

```
##           id           rx           sex           age           obstruct
##  Min.      : 1      Obs       :315      Min.      :0.000      Min.      :18.00      Min.      :0.0000
## 1st Qu.:233      Lev       :310      1st Qu.:0.000      1st Qu.:53.00      1st Qu.:0.0000
## Median :465      Lev+5FU:304      Median :1.000      Median :61.00      Median :0.0000
## Mean    :465                                Mean    :0.521      Mean    :59.75      Mean    :0.1938
## 3rd Qu.:697                                3rd Qu.:1.000      3rd Qu.:69.00      3rd Qu.:0.0000
## Max.    :929                                Max.     :1.000      Max.     :85.00      Max.     :1.0000
##
##           perfor           adhere           nodes           differ
##  Min.      :0.00000      Min.      :0.0000      Min.      : 0.00      Min.      :1.000
## 1st Qu.:0.00000      1st Qu.:0.0000      1st Qu.: 1.00      1st Qu.:2.000
## Median :0.00000      Median :0.0000      Median : 2.00      Median :2.000
## Mean    :0.02906      Mean    :0.1453      Mean    : 3.66      Mean    :2.063
## 3rd Qu.:0.00000      3rd Qu.:0.0000      3rd Qu.: 5.00      3rd Qu.:2.000
## Max.    :1.00000      Max.     :1.0000      Max.     :33.00      Max.     :3.000
##                               NA's      :18      NA's      :23
##           extent           surg           node4           time
##  Min.      :1.000      Min.      :0.0000      Min.      :0.0000      Min.      : 8
## 1st Qu.:3.000      1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.: 370
## Median :3.000      Median :0.0000      Median :0.0000      Median :1548
## Mean    :2.887      Mean    :0.2659      Mean    :0.2745      Mean    :1405
## 3rd Qu.:3.000      3rd Qu.:1.0000      3rd Qu.:1.0000      3rd Qu.:2289
## Max.    :4.000      Max.     :1.0000      Max.     :1.0000      Max.     :3329
```

```
##
##      status      timeD      statusD      timeR
##  Min.   :0.0000  Min.    : 23  Min.    :0.0000  Min.    : 8
##  1st Qu.:0.0000  1st Qu.: 806  1st Qu.:0.0000  1st Qu.: 370
##  Median :1.0000  Median :1976  Median :0.0000  Median :1548
##  Mean   :0.5856  Mean   :1670  Mean   :0.4865  Mean   :1405
##  3rd Qu.:1.0000  3rd Qu.:2364  3rd Qu.:1.0000  3rd Qu.:2289
##  Max.   :2.0000  Max.    :3329  Max.    :1.0000  Max.    :3329
##
##      statusR
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :1.0000
##  Mean   :0.5038
##  3rd Qu.:1.0000
##  Max.   :1.0000
##
```

Question 2

We keep only the observations corresponding to patients of these two groups.

```
d <- d[which(d$rx %in% c("Obs", "Lev+5FU")),]
d$rx <- droplevels(d$rx)  # drop the level no longer present in the dataset
```

We now look at the number of patients included in each treatment group.

```
table(d$rx)
```

```
##
##      Obs Lev+5FU
##      315      304
```

We read that 315 did not receive the treatment and 304 did.

Question 3

Before creating the table, we first create factor variables for all the categorical variables and use appropriate labels.

```
d$sex <- factor(d$sex, levels=c(1,0), labels=c("male", "female"))
d$obstruct <- factor(d$obstruct, levels=c(0,1), labels=c("no", "yes"))
d$perfor <- factor(d$perfor, levels=c(0,1), labels=c("no", "yes"))
d$adhere <- factor(d$adhere, levels=c(0,1), labels=c("no", "yes"))
d$differ <- factor(d$differ, levels=c(1,2,3), labels=c("well", "moderate", "poor"))
```

```
d$extent <- factor(d$extent,levels=c(1,2,3,4),labels=c("submucosa",
                                                    "muscle",
                                                    "serosa",
                                                    "contiguous structures"))
d$surg <- factor(d$surg,levels=c(0,1),labels=c("short","long"))
d$node4 <- factor(d$node4,levels=c(0,1),labels=c("<=4",">4"))
```

We are now better prepared to create a descriptive table. The aim of this table is to summarize the baseline variable distributions in each treatment group. We report median, first and third quartile for quantitative variables and frequencies and percentages for categorical variables. As recommended by guidelines (e.g. CONSORT), we do not present useless p-values to test for differences between the two groups.

```
library(Publish)
tab1 <- univariateTable(rx~Q(age) + Q(nodes) + sex +
                        obstruct + perfor + adhere +
                        differ + extent + surg + node4,
                        data=d,
                        compare.groups = FALSE,
                        show.totals = FALSE)
tab1
```

##	Variable	Level	Obs (n=315)	Lev+5FU (n=304)
## 1	age	median [iqr]	60 [53, 68]	62 [52, 70]
## 2	nodes	median [iqr]	2 [1, 5]	2 [1, 4]
## 3		missing	3	9
## 4	sex	male	166 (52.7)	141 (46.4)
## 5		female	149 (47.3)	163 (53.6)
## 6	obstruct	no	252 (80.0)	250 (82.2)
## 7		yes	63 (20.0)	54 (17.8)
## 8	perfor	no	306 (97.1)	296 (97.4)
## 9		yes	9 (2.9)	8 (2.6)
## 10	adhere	no	268 (85.1)	265 (87.2)
## 11		yes	47 (14.9)	39 (12.8)
## 12	differ	well	27 (8.8)	29 (9.7)
## 13		moderate	229 (74.4)	215 (72.1)
## 14		poor	52 (16.9)	54 (18.1)
## 15		missing	7	6
## 16	extent	submucosa	8 (2.5)	10 (3.3)
## 17		muscle	38 (12.1)	32 (10.5)
## 18		serosa	249 (79.0)	251 (82.6)
## 19		contiguous structures	20 (6.3)	11 (3.6)
## 20	surg	short	224 (71.1)	228 (75.0)
## 21		long	91 (28.9)	76 (25.0)
## 22	node4	<=4	228 (72.4)	225 (74.0)

23

>4 87 (27.6) 79 (26.0)

We see that the distributions of all the covariates seem to be fairly similar in the two groups. This is not at all a surprise: this was expected because of the randomization of the treatment allocation. Because of the random allocation to each treatment group, the patients are as old/young and healthy/commorbid in the two groups. The two populations differ only by treatment. Hence, if we observe a survival difference when comparing the two groups (at a later question), we can confidently conclude that this must be a consequence of the treatment.

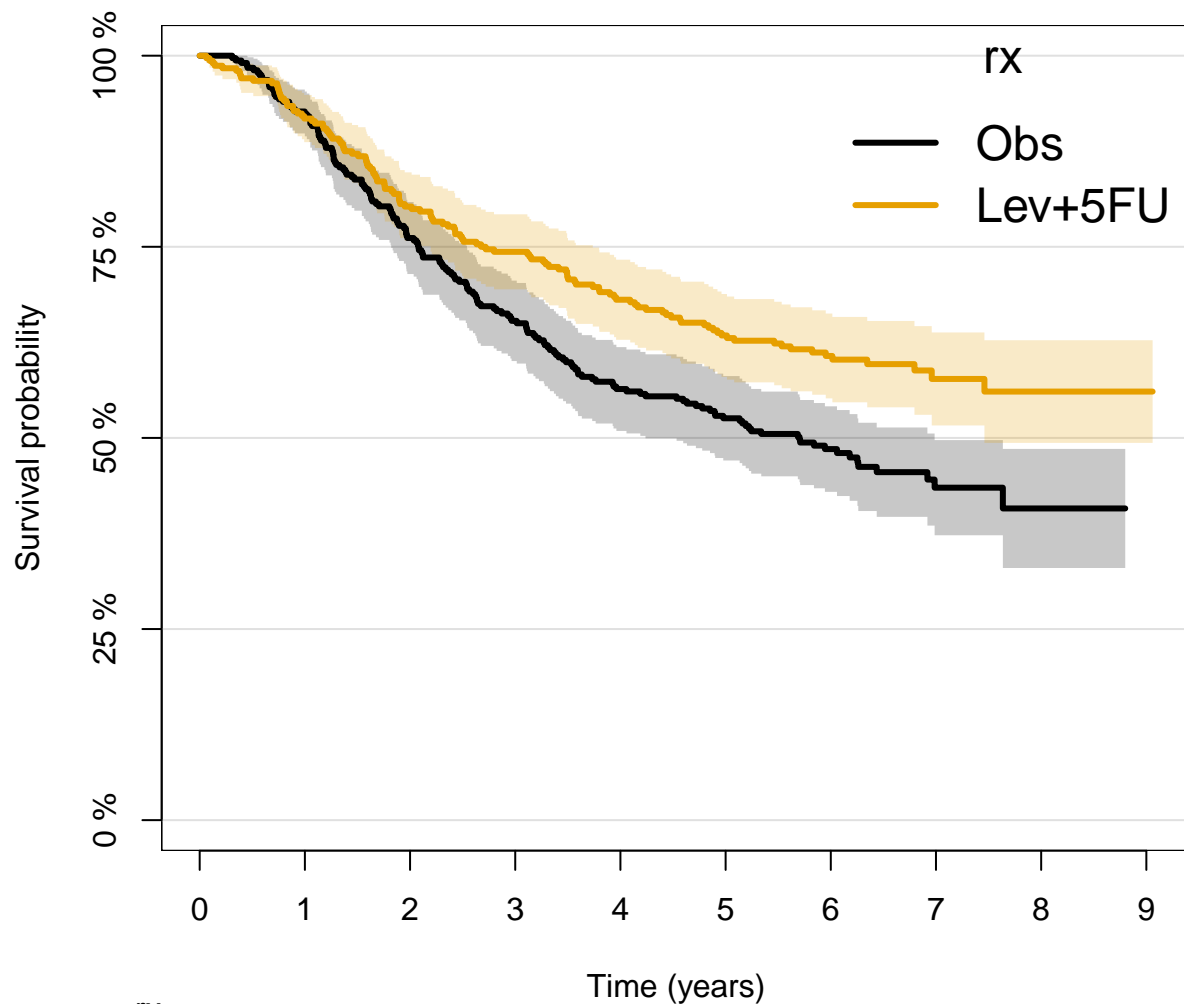
Question 4

We now use the Kaplan-Meier method to estimate the survival curves in each treatment group. We use the `timeD` and `statusD` variables, not `time` and `status`. This is because according to the description of the data, they are those relevant when studying all-cause death as the endpoint.

```
library(prodlim)
KM1 <- prodlim(Surv(timeD,statusD)~rx,data=d)
```

We can then plot the curves. We use the option `timeconverter="days2years"` to show the years instead of days. The time unit is “days” in the dataset, but the interpretation is easier using “years”. We could of course further fine tune the plot (if relevant) as exemplified in the R-demo.

```
plot(KM1,timeconverter="days2years")
```



rx		Time (years)									
Obs:		315	291	239	205	177	160	101	41	7	0
Lev+5FU:		304	279	244	226	205	187	128	52	12	2

The plot suggests that the treatment improves the survival chances of the patients. Indeed, the survival curve corresponding to those treated is well above that of those not treated. This is the case everywhere (i.e. for all time points) maybe except at the very beginning (say within the first year), but there is anyway too much uncertainty there (as shown by the confidence intervals) to draw a clear conclusion. The (pointwise) 95% confidence intervals do not really cross after approximately 3 years. This suggests that the data contain rather strong evidence that the treatment work.

Question 5

We now compute the median survival times (with 95%-CI) for each treatment group.

```
quantile(KM1)
```

```
## Quantiles of the event time distribution based on the method.
```

```
##
## Table of quantiles and corresponding confidence limits:
##
## rx=Obs
##
##      q quantile lower upper
## 1 0.00      NA      NA      NA
## 2 0.25      NA      NA      NA
## 3 0.50    2083    1548    2552
## 4 0.75     760     665     928
## 5 1.00     113     113     164
## Median time (IQR):2083.00 (760.00;--)
```

```
##
## Table of quantiles and corresponding confidence limits:
##
## rx=Lev+5FU
##
##      q quantile lower upper
## 1 0.00      NA      NA      NA
## 2 0.25      NA      NA      NA
## 3 0.50      NA    2725      NA
## 4 0.75     993     802    1365
## 5 1.00      23      23      52
## Median time (IQR):-- (993.00;--)
```

We estimate that, without treatment, half of the patients die within 2083 days (95%-CI=[1548, 2552]). We can convert to years as follows.

```
round(c(2083, 1548, 2552)/365,1)
```

```
## [1] 5.7 4.2 7.0
```

That is 5.7 years (95%-CI=[4.2,7.0]). We could have read (approximately) this results from the Kaplan-Meier plot we previously produced. Indeed, we just need to read the x-values (i.e., the times) at which the survival curve and 95%-CI intersect the $y=0.5$ horizontal line (see lecture slides).

For the treated, we do not obtain an estimated value for the median. We obtain a value for the lower limit of the confidence interval only (2725 days, i.e. 7.5 years). Therefore, we can only conclude that we are 95% confident that the median survival time is larger than 7.5 years. This is because only the curve of the lower limit of the (pointwise) confidence interval intersect with the $y=0.5$ horizontal line. We do not observe enough deaths to know more about the median survival time for this group.

Question 6

We now report the estimated 7-year survival probabilities in each group, together with a 95%-CI.

```
KM10 <- summary(KM1,time=7*365)$table$`rx=Obs`  
KM10 # results for group rx=0  
  
##   time n.risk n.event n.lost      surv   se.surv   lower   upper  
## 1 2555     41       0       0 0.4349148 0.03177142 0.372644 0.4971857
```

```
KM11 <- summary(KM1,time=7*365)$table$`rx=Lev+5FU`  
KM11 # results for group rx=1  
  
##   time n.risk n.event n.lost      surv   se.surv   lower   upper  
## 1 2555     52       0       0 0.5771258 0.0310435 0.5162816 0.6379699
```

We read that we estimated the 7-year survival probability to be 43.5% (95%-CI=[37.3,49.7]) for those who did not receive the treatment, versus 57.7% (95%-CI=[51.6,63.8]) for those who did.

We now report the estimated difference in 7-year survival probability, together with a 95%-CI and a p-value.

```
# Second, compute the difference  
diffSurv <- KM11[1,"surv"] - KM10[1,"surv"]  
# Third, compute the s.e. of the difference  
seDiffSurv <- sqrt(KM11[1,"se.surv"]^2 + KM10[1,"se.surv"]^2)  
# Now compute the 95% CI  
lowerDiffSurv <- diffSurv - qnorm(1-0.05/2)*seDiffSurv  
upperDiffSurv <- diffSurv + qnorm(1-0.05/2)*seDiffSurv  
# And the -value  
pvalDiffSurv <- 2*(1-pnorm(abs(diffSurv/seDiffSurv)))  
# Put all the results together  
ResDiffSurv <- c(Est=diffSurv,  
                 lower=lowerDiffSurv,  
                 upper=upperDiffSurv,  
                 p=pvalDiffSurv)  
# print the difference, 95% CI and p-value  
round(ResDiffSurv,3)
```

```
##   Est lower upper    p  
## 0.142 0.055 0.229 0.001
```

We read that we estimate a statistically significant difference of 14.2% (95%-CI=[5.5,22.9], $p=0.001$) in favor of those receiving the treatment. By “statistically significant”, we mean that the data contain enough evidence to reject the null hypothesis of no difference in 7-year survival probability.

Question 7

Before we move on to the main analysis (at question 11), we first compare the survival curves of the two treatment groups via a simple log-rank test (just to practice and explore the data further).

```
library(survival)
ResLogRank <- survdiff(Surv(timeD,statusD)~rx,data=d)
print(ResLogRank,digits=6)

## Call:
## survdiff(formula = Surv(timeD, statusD) ~ rx, data = d)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## rx=Obs       315      168  141.117    5.12134    9.96567
## rx=Lev+5FU  304      123  149.883    4.82180    9.96567
##
##  Chisq= 10  on 1 degrees of freedom, p= 0.001595
```

We read that the p-value is significant ($p=0.002$). The null hypothesis of this test is that the survival curves of the two groups are identical (everywhere, i.e. at all times). The small p-value is not surprising, as we have previously seen that the curves are estimated to be very different, with (pointwise) 95%-CI which mostly do not overlap for time points between say 3 and 9 years (approximately).

Question 8

To accompany the p-value of the log-rank test, it is considered good practice to report an estimated “effect size” and 95%-CI. That is, the hazard ratio obtained from a univariate Cox model and its 95%-CI. We use the `coxph` function to compute this as follows.

```
cox1 <- coxph(Surv(timeD,statusD)~rx,data=d)
summary(cox1)

## Call:
## coxph(formula = Surv(timeD, statusD) ~ rx, data = d)
##
##    n= 619, number of events= 291
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## rxLev+5FU -0.3728    0.6888   0.1188 -3.138   0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## rxLev+5FU    0.6888      1.452    0.5457    0.8694
```



```
##
## Concordance= 0.542 (se = 0.015 )
## Likelihood ratio test= 9.98 on 1 df, p=0.002
## Wald test = 9.85 on 1 df, p=0.002
## Score (logrank) test = 9.96 on 1 df, p=0.002
```

We read that the hazard ratio is estimated to 0.69 (95%-CI=[0.55,0.87], p=0.002). Here the p-value corresponds to the null hypothesis of a hazard ratio equal to 1 (i.e. same hazard rate in both groups). We can therefore conclude that the hazard rate (i.e. the instantaneous risk of death among patients still alive at any time during the follow-up) is 0.69 times smaller among the treated patients than among the patients who did not receive the treatment (95%-CI=[0.55,0.87], p=0.002).

We can also report the results in terms of a relative *instantaneous risk* (or rate) reduction, by computing “one minus the hazard ratio”.

```
round((1-exp(cox1$coef))*100,1)
```

```
## rxLev+5FU
## 31.1
```

```
round(rev(1-exp(confint(cox1)))*100,1)
```

```
## [1] 13.1 45.4
```

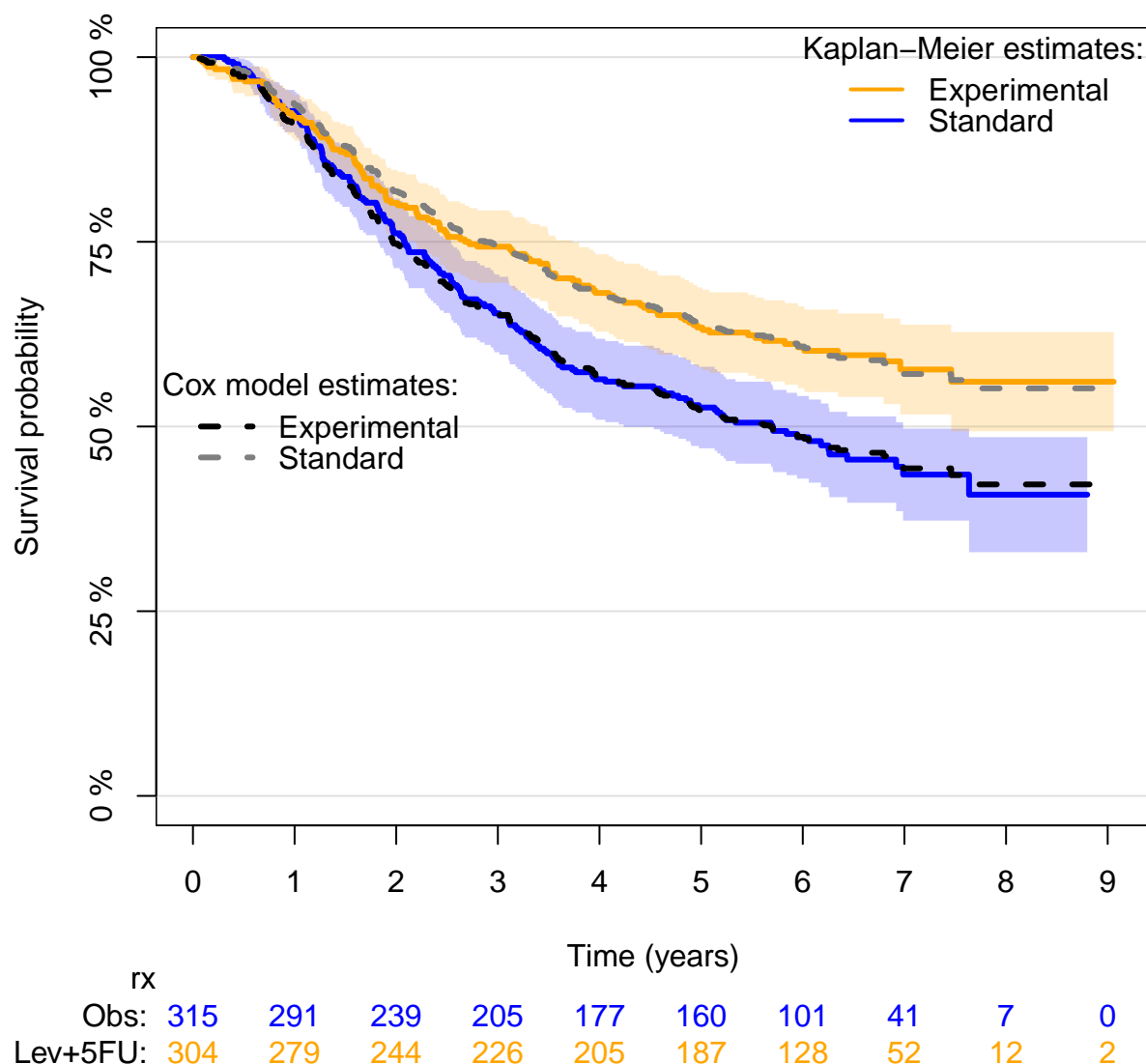
Hence, we conclude that patients receiving fluorouracil plus levamisole had a 31.1% reduction in mortality rate (95%-CI=[13.1%,45.4%]). These results approximately match those of Moertel et al (1995), who analyzed an almost identical version of these data and found 33% (95%-CI=[16%,47%]).

Question 9

We now produce a simple plot to visually compare the survival curves for each treatment group, when they are estimated via either Kaplan-Meier or a univariate Cox model. This is to check the proportional hazards assumption of the Cox model.

```
# Step 1: create "new data" to predict the survival
dnew1 <- data.frame(rx="Obs")
dnew0 <- data.frame(rx="Lev+5FU")
# Step 2: predict/estimate the survival probabilities for both groups
# using the Cox model
scox1 <- survfit(cox1,newdata=dnew1)
scox0 <- survfit(cox1,newdata=dnew0)
# Step 3: plot Kaplan-Meier curves
plot(KM1,legend=FALSE,col=c("blue","orange"),timeconverter="days2years")
# Step 4: add the estimated survival curves from the Cox model
lines(scox0$time,scox0$surv,col="grey50",type="s",lwd=3,lty=2)
```

```
lines(scox1$time,scox1$surv,col="black",type="s",lwd=3,lty=2)
# Step 5: add legends
legend("left",legend=c("Experimental","Standard"),lwd=3,lty=2,
      col=c("black","grey50"),title="Cox model estimates: ",bty="n")
legend("topright",legend=c("Experimental","Standard"),lwd=2,lty=1,
      col=c("orange","blue"),title="Kaplan-Meier estimates:",bty="n")
```



The curves estimated via both methods are very similar. We therefore conclude that nothing seems wrong with the proportional hazards assumption of the Cox model. Hence, we can further conclude that the hazard ratio that we have estimated via the Cox model is a sensible “summary” of the difference between the two survival curves. Had the model fit been bad, this would not have been a good “summary” of the difference.

Reminder: The Kaplan-Meier method does not make any assumption about how the curves differ from each other. By contrast, the Cox model does make such an assumption, via the proportional hazards assumption. For instance, the Cox model assumes that the two curves

cannot cross (which is a consequence of the proportional hazards assumption).

Question 10

Before we move on to the main analysis (at question 11), we first compare the restricted mean survival times (RMST) at $t = 7$ years for the two treatment groups (just to practice and explore the data further).

We first need to create a 0/1 binary variable for the treatment group to use the `rmst2` function of the package `survRM2`.

```
d$trt <- as.numeric(d$rx=="Lev+5FU")
table(d$trt)
```

```
##
##    0    1
## 315 304
```

We can now use the `rmst2` function as follows.

```
library(survRM2)
ResRMST <- rmst2(time=d$timeD/365, # trick to have a time unit in years
                 status=d$statusD,
                 arm=d$trt,
                 tau=7)           # Beware of the time unit!
ResRMST
```

```
##
## The truncation time: tau = 7  was specified.
##
## Restricted Mean Survival Time (RMST) by arm
##           Est.    se lower .95 upper .95
## RMST (arm=1) 5.190 0.14    4.917    5.464
## RMST (arm=0) 4.635 0.14    4.360    4.910
##
##
## Restricted Mean Time Lost (RMTL) by arm
##           Est.    se lower .95 upper .95
## RMTL (arm=1) 1.810 0.14    1.536    2.083
## RMTL (arm=0) 2.365 0.14    2.090    2.640
##
##
## Between-group contrast
##           Est. lower .95 upper .95    p
## RMST (arm=1)-(arm=0) 0.556    0.168    0.944 0.005
## RMST (arm=1)/(arm=0) 1.120    1.034    1.212 0.005
```

```
## RMTL (arm=1)/(arm=0) 0.765      0.632      0.926 0.006
```

We estimate that, if we treat future patients from the same population with fluorouracil plus levamisole and follow them for 7 years, the average time spent alive will be approximately 5.2 years (95%-CI=[4.9,5.5]). By contrast, if we do not treat future patients from the same population and follow them for 7 years, the average time spent alive will be approximately 4.6 years (95%-CI=[4.4,4.9]). That is, 0.56 years less (95%-CI=[0.17,0.94], $p=0.005$).

Question 11

We now proceed to the main analysis. We fit a multiple Cox regression model to compare the survival chances of a patient who received levamisole plus fluorouracil to that of a patient who did not, when both patients are similar with regards to:

- obstruction of the colon by the tumor (**obstruct**),
- number of positive lymph nodes (more or less than 4, **node4**)
- time from surgery to inclusion into the trial (short vs long, **surg**)
- age and sex

We assume that the following was prespecified. We do not model any interaction and age is used in the model as a continuous variable with a (usual/simple) log-linear effect on the hazard rate.

```
cox2 <- coxph(Surv(timeD,statusD)~rx+obstruct+node4+surg+age+sex,data=d)
summary(cox2)
```

```
## Call:
## coxph(formula = Surv(timeD, statusD) ~ rx + obstruct + node4 +
##      surg + age + sex, data = d)
##
##      n= 619, number of events= 291
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## rxLev+5FU    -0.360768  0.697141  0.119597 -3.017  0.00256 **
## obstructyes   0.152375  1.164597  0.146792  1.038  0.29926
## node4>4       0.937406  2.553348  0.121789  7.697 1.39e-14 ***
## surglong      0.241523  1.273187  0.127573  1.893  0.05833 .
## age           0.003651  1.003658  0.004848  0.753  0.45140
## sexfemale     0.090796  1.095046  0.118013  0.769  0.44167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## rxLev+5FU          0.6971      1.4344    0.5515    0.8813
```

```
## obstructyes      1.1646      0.8587      0.8734      1.5528
## node4>4          2.5533      0.3916      2.0111      3.2417
## surglong         1.2732      0.7854      0.9915      1.6349
## age              1.0037      0.9964      0.9942      1.0132
## sexfemale        1.0950      0.9132      0.8689      1.3800
##
## Concordance= 0.646 (se = 0.016 )
## Likelihood ratio test= 69.76 on 6 df, p=5e-13
## Wald test          = 75.73 on 6 df, p=3e-14
## Score (logrank) test = 80.09 on 6 df, p=3e-15
```

We estimate the hazard ratio to be 0.70 (95%-CI=[0.55,0.88], $p=0.003$). That is, we estimate that the mortality rate is reduced by 30% (95%-CI=[12%,45%], $p=0.003$), for a treated patient, as compared to a patient who is not treated, when both have the same age and sex, and are also similar with respect to time from surgery to inclusion into the trial, number of positive lymph nodes and obstruction of the colon by the tumor. The main analysis shows significant results: we obtained a positive finding!

Question 12

One could wonder whether modeling a log-linear effect of age on the hazard rate was a good idea. It is, after all and to some extent, an arbitrary choice. It can be interesting to compare the previous conclusions to those obtained when modeling the effect of age via age groups, which does not require to assume a log-linear effect. For instance, we can use four age groups: 18-50, 50-60, 60-70, 70-85.

```
d$agec <- cut(d$age,breaks=c(18,50,60,70,85),include.lowest=TRUE)
table(d$agec,useNA="always")
```

```
##
## [18,50] [50,60] [60,70] [70,85] <NA>
##      132      167      202      118      0
```

```
cox2c <- coxph(Surv(timeD,statusD)~rx+obstruct+node4+surg+agec+sex,data=d)
summary(cox2c)
```

```
## Call:
## coxph(formula = Surv(timeD, statusD) ~ rx + obstruct + node4 +
##      surg + agec + sex, data = d)
##
## n= 619, number of events= 291
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## rxLev+5FU    -0.38714   0.67899  0.12043 -3.215  0.00131 **
## obstructyes   0.12979   1.13859  0.14685  0.884  0.37678
## node4>4       0.93300   2.54212  0.12179  7.661 1.85e-14 ***
```

```
## surglong      0.23194    1.26104    0.12759    1.818    0.06909 .
## agec(50,60] -0.21717    0.80479    0.17348   -1.252    0.21062
## agec(60,70]  0.14804    1.15956    0.15892    0.932    0.35157
## agec(70,85] -0.03359    0.96697    0.18601   -0.181    0.85672
## sexfemale    0.10476    1.11045    0.11871    0.883    0.37750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## rxLev+5FU      0.6790      1.4728    0.5362    0.8598
## obstructyes    1.1386      0.8783    0.8538    1.5183
## node4>4        2.5421      0.3934    2.0023    3.2275
## surglong       1.2610      0.7930    0.9820    1.6193
## agec(50,60]    0.8048      1.2426    0.5728    1.1307
## agec(60,70]    1.1596      0.8624    0.8492    1.5833
## agec(70,85]    0.9670      1.0342    0.6716    1.3924
## sexfemale      1.1104      0.9005    0.8799    1.4014
##
## Concordance= 0.65 (se = 0.016 )
## Likelihood ratio test= 74.72 on 8 df,  p=6e-13
## Wald test              = 80.38 on 8 df,  p=4e-14
## Score (logrank) test = 84.74 on 8 df,  p=5e-15
```

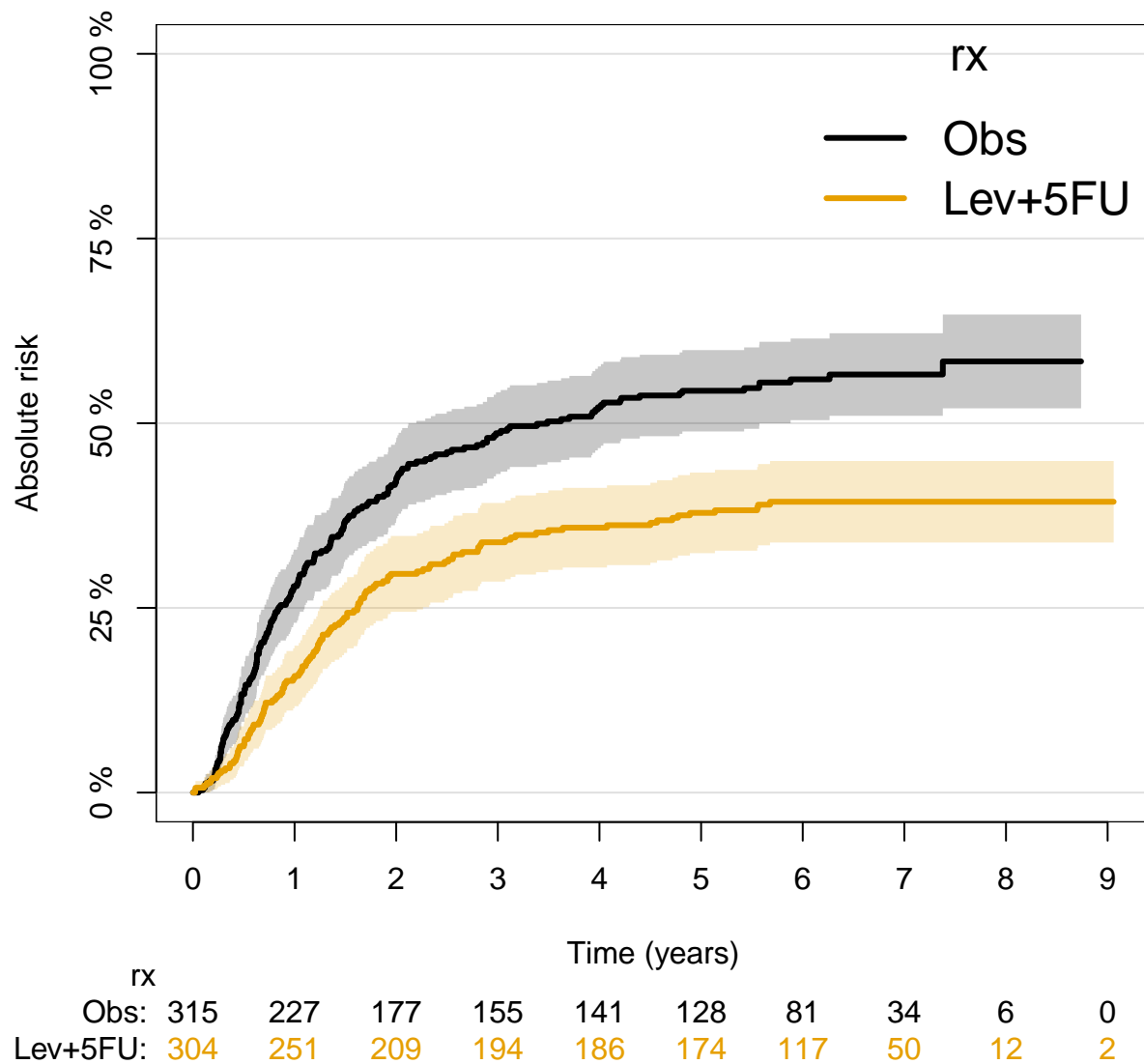
We now estimate the hazard ratio to be 0.68 (95%-CI=[0.54,0.86], p=0.001). The main conclusion does not change at all. This simple “sensitivity analysis” is very reassuring.

Exercise B

(only the R code is provided)

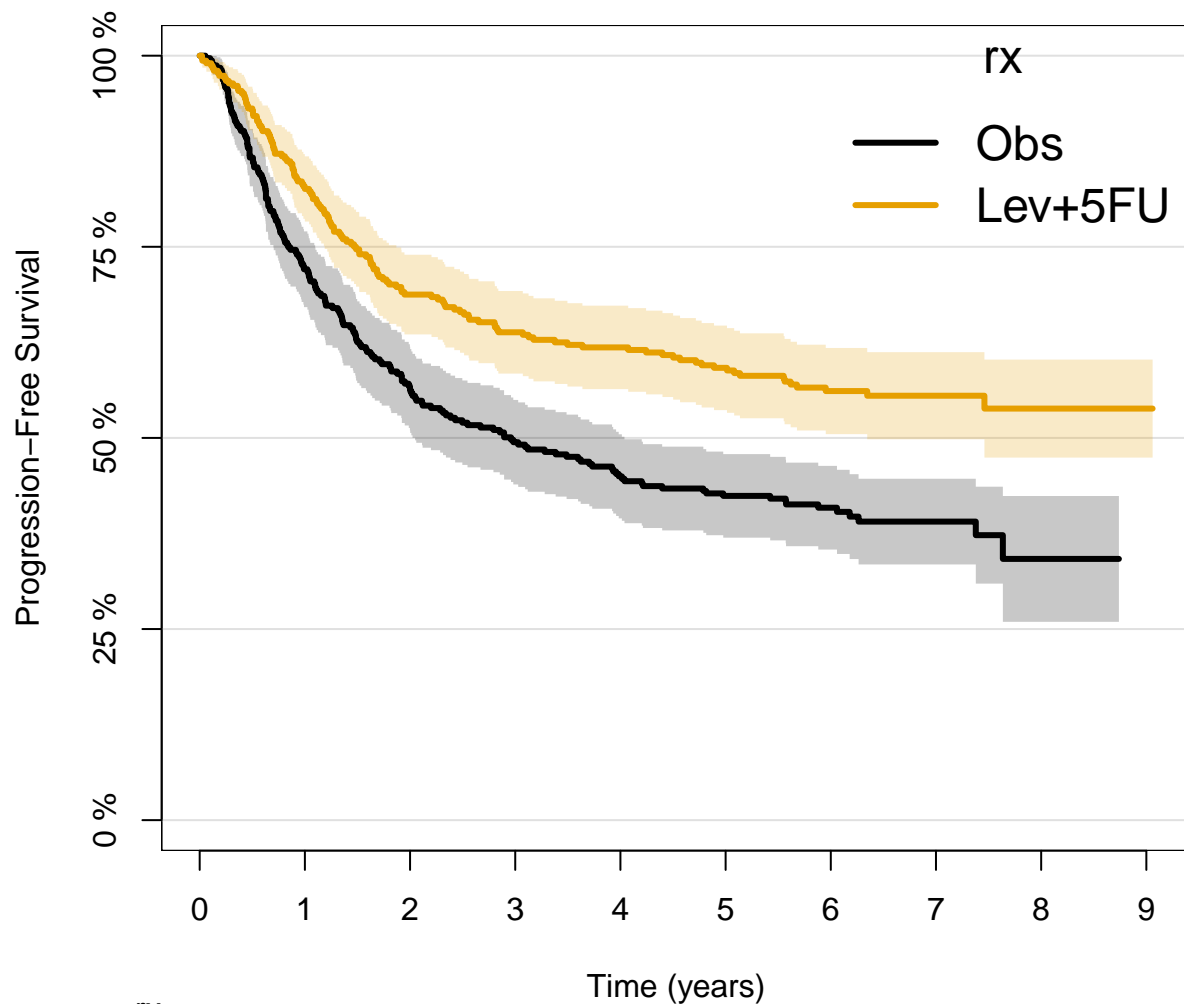
Question 1

```
AJ1 <- prodlim(Hist(time,status)~rx,data=d)
plot(AJ1,timeconverter="days2years")
```



Question 2

```
KMPFS <- prodlim(Surv(time,status!=0)~rx,data=d)
plot(KMPFS,timeconverter="days2years",ylab="Progression-Free Survival")
```



rx										
Obs:	315	227	177	155	141	128	81	34	6	0
Lev+5FU:	304	251	209	194	186	174	117	50	12	2

```
cox2PFS <- coxph(Surv(time,status!=0)~rx+obstruct+node4+surg+age+sex,data=d)
summary(cox2PFS)
```

```
## Call:
## coxph(formula = Surv(time, status != 0) ~ rx + obstruct + node4 +
##      surg + age + sex, data = d)
##
##      n= 619, number of events= 324
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## rxLev+5FU -0.4605591  0.6309308  0.1136788 -4.051 5.09e-05 ***
## obstructyes  0.1249044  1.1330401  0.1401069  0.891  0.3727
## node4>4      0.8466340  2.3317849  0.1162072  7.286 3.20e-13 ***
## surglong     0.2492774  1.2830979  0.1206781  2.066  0.0389 *
## age         -0.0003737  0.9996264  0.0046560 -0.080  0.9360
## sexfemale    0.1134394  1.1201240  0.1118273  1.014  0.3104
```



```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## rxLev+5FU      0.6309      1.5850      0.5049      0.7884
## obstructyes    1.1330      0.8826      0.8610      1.4911
## node4>4        2.3318      0.4289      1.8568      2.9282
## surglong       1.2831      0.7794      1.0128      1.6255
## age            0.9996      1.0004      0.9905      1.0088
## sexfemale      1.1201      0.8928      0.8997      1.3946
##
## Concordance= 0.647 (se = 0.015 )
## Likelihood ratio test= 73.57 on 6 df,  p=8e-14
## Wald test              = 78.85 on 6 df,  p=6e-15
## Score (logrank) test = 82.5 on 6 df,  p=1e-15

```