Stat 525 Homework 10

Kenny Flagg

November 21, 2016

The data are of survival times of patients with acute myelogenous leukemia (AML). The question at the time was whether the standard course of chemotherapy should be extended ('maintainance') for additional cycles. The data can be found in the data file ***aml*** *in the* ***survival*** *package. We will look at Kaplan-Meier and Nelson-Aalen nonparametric survival curves and carry out the log-rank test to compare the patients in the two groups.*

1. *Print out the data set just so you have it in front of you. It is not that big:* ***time*** *is survival time,* ***status*** *is censoring status with 1 denoting death and 0 denoting censored and* ***x*** *is a binary categorical variable indicating membership in either the Maintained or Non-maintained groups.*

```
library(survival)
aml
```

```
   time status             x
1     9      1    Maintained
2    13      1    Maintained
3    13      0    Maintained
4    18      1    Maintained
5    23      1    Maintained
6    28      0    Maintained
7    31      1    Maintained
8    34      1    Maintained
9    45      0    Maintained
10   48      1    Maintained
11  161      0    Maintained
12    5      1 Nonmaintained
13    5      1 Nonmaintained
14    8      1 Nonmaintained
15    8      1 Nonmaintained
16   12      1 Nonmaintained
17   16      0 Nonmaintained
18   23      1 Nonmaintained
19   27      1 Nonmaintained
20   30      1 Nonmaintained
21   33      1 Nonmaintained
22   43      1 Nonmaintained
23   45      1 Nonmaintained
```

2. *The survival times for the Maintained group is*

```
9     13     13+    18    23    28+    31    34    45+    48    161+
```

Fill in the table below by hand. Remember $\widehat{P}_k$ is actually the survival function estimate.

time	$\widehat{q}_i$	$\widehat{p}_i$	$\widehat{P}_k$
9	1/11	10/11	$10/11 = 0.909$
13	1/10	9/10	$0.909 \times 9/10 = 0.818$
18	1/8	7/8	$0.818 \times 7/8 = 0.716$
23	1/7	6/7	$0.716 \times 6/7 = 0.614$
31	1/5	4/5	$0.614 \times 4/5 = 0.491$
34	1/4	3/4	$0.491 \times 3/4 = 0.368$
48	1/2	1/2	$0.368 \times 1/2 = 0.184$

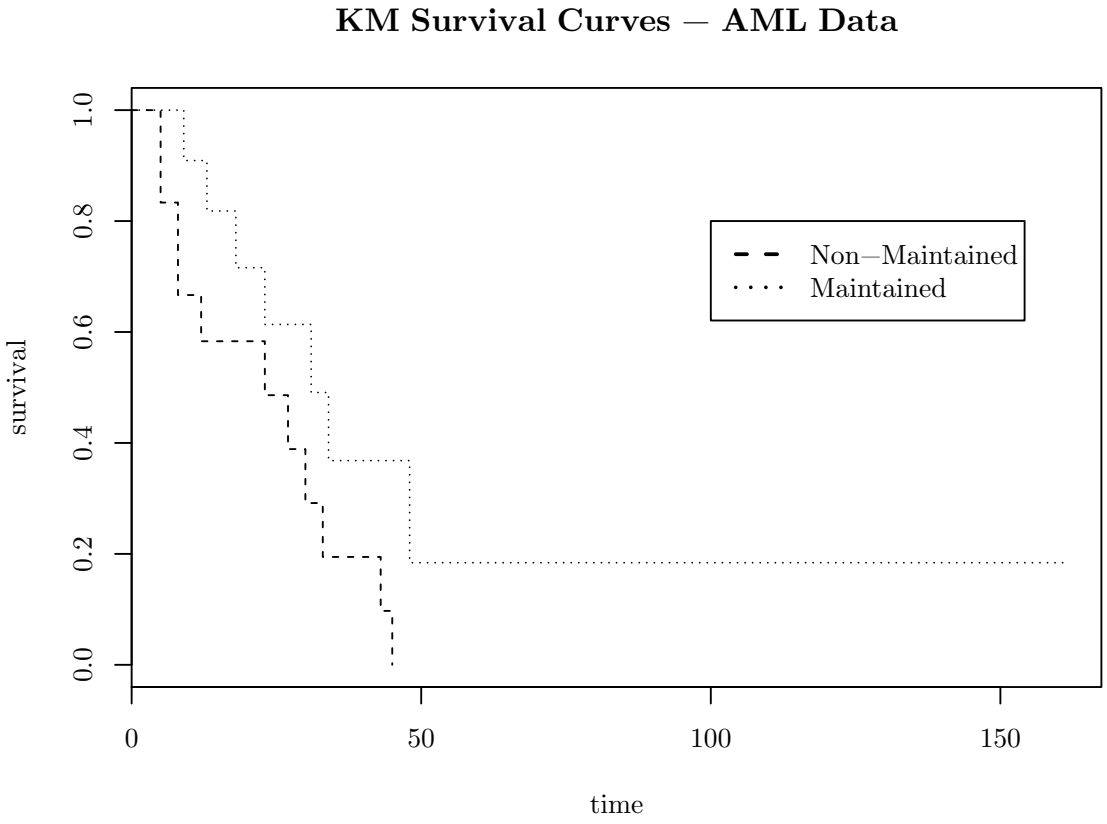
3. *Give summary information including mean and median survival times, and Kaplan-Meier Curves for the two groups. Briefly discuss what these results seem to indicate about the two groups.*

For all times after time 5, the estimated survival curve for the Maintained group is above the estimated survival curve for the Non-maintained group. This indicates that individuals whose chemotherapy is maintained are more likely to survive to a given time than individuals in the whose chemotherapy is not maintained.

```
aml.obj <- Surv(aml$time, aml$status)
fitKM <- survfit(aml.obj ~ x, data = aml, conf.type = 'log-log')
print(fitKM, print.rmean = TRUE)

Call: survfit(formula = aml.obj ~ x, data = aml, conf.type = "log-log")

                 n events *rmean *se(rmean) median 0.95LCL 0.95UCL
x=Maintained    11      7   42.0      11.26     31      13      NA
x=Nonmaintained 12     11   22.7       4.18     23       5      33
    * restricted mean with upper limit =  103

summary(fitKM)

Call: survfit(formula = aml.obj ~ x, data = aml, conf.type = "log-log")

                x=Maintained
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    9     11       1    0.909  0.0867       0.5081        0.987
   13     10       1    0.818  0.1163       0.4474        0.951
   18      8       1    0.716  0.1397       0.3502        0.899
   23      7       1    0.614  0.1526       0.2658        0.835
   31      5       1    0.491  0.1642       0.1673        0.753
   34      4       1    0.368  0.1627       0.0928        0.657
   48      2       1    0.184  0.1535       0.0117        0.525

                x=Nonmaintained
```

```
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    5     12       2   0.8333  0.1076      0.48171        0.956
    8     10       2   0.6667  0.1361      0.33702        0.860
   12      8       1   0.5833  0.1423      0.27014        0.801
   23      6       1   0.4861  0.1481      0.19188        0.730
   27      5       1   0.3889  0.1470      0.12627        0.650
   30      4       1   0.2917  0.1387      0.07240        0.561
   33      3       1   0.1944  0.1219      0.03120        0.461
   43      2       1   0.0972  0.0919      0.00575        0.349
   45      1       1   0.0000     NaN           NA           NA
```

```
plot(fitKM, lty = 3:2, xlab = 'time', ylab = 'survival')
legend(100, 0.8, c('Non-Maintained', 'Maintained'), lty = 2:3, lwd = 2)
title(main = 'KM Survival Curves - AML Data')
```



4. *Compare the mean survival times for the two groups and the median survival times for the two groups. Which do you think is more important to use as a summary measure for comparison - the mean or the median? Justify your answer.*

   The mean survival time for the Maintained group is 42.0, much larger than the mean of 22.7 for the Non-maintained group. The medians are closer together than the means are, with the median being 31 for the Maintained group and 23 for the Non-maintained group. The median is more useful because the mean for the Maintained group is probably inflated by the one individual who was still alive at time 161.

5. *Carry out the log-rank test and give the results. What do they imply about the survival times of the two groups. Justify your answer.*

```
survdiff(Surv(time, status) ~ x, aml)

Call:
survdiff(formula = Surv(time, status) ~ x, data = aml)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
x=Maintained    11        7    10.69      1.27       3.4
x=Nonmaintained 12       11     7.31      1.86       3.4

 Chisq= 3.4  on 1 degrees of freedom, p= 0.0653
```

   The log-rank statistic is $\chi^2_1 = 3.4$ with p-value $= 0.0653$, providing moderate evidence that the Maintained and Non-maintained groups have different survival functions.

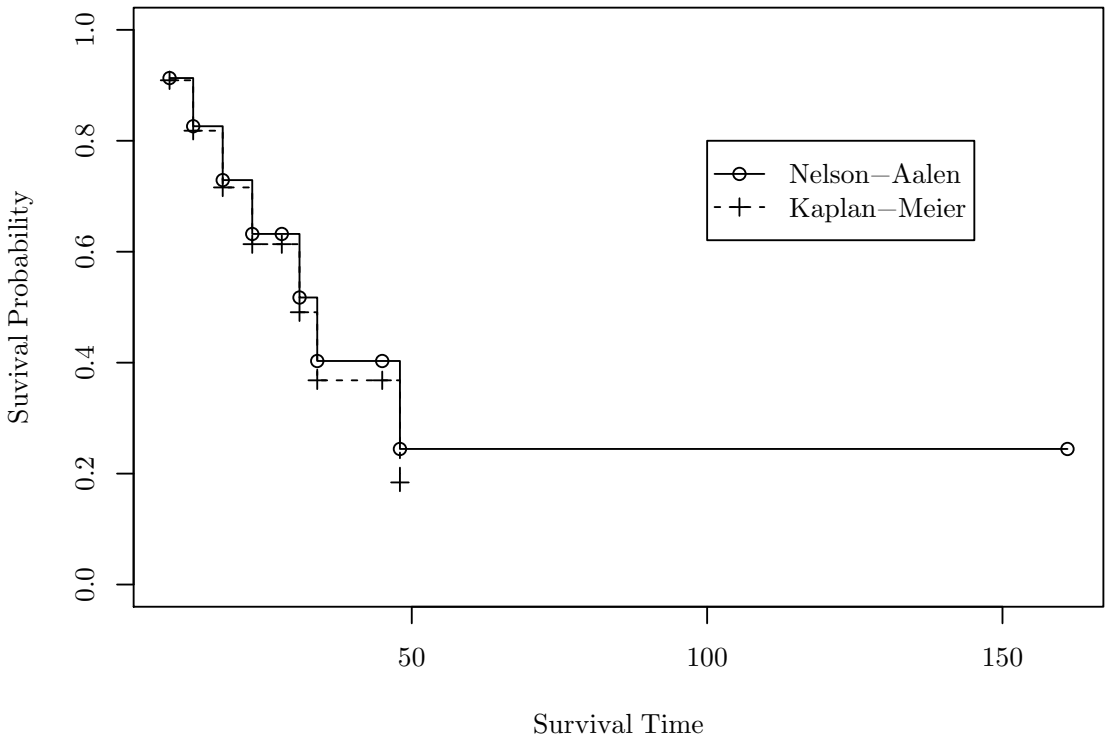
6. *Using the data in the table in Problem 2 calculate the Nelson-Aalen estimator of the survival curve.*

time	$\widehat{q}_i$	$\widehat{H}$	$\exp(-\widehat{H})$
9	$1/11$	$1/11 = 0.091$	$\exp(-0.091) = 0.913$
13	$1/10$	$0.091 + 1/10 = 0.191$	$\exp(-0.191) = 0.826$
18	$1/8$	$0.191 + 1/8 = 0.316$	$\exp(-0.316) = 0.729$
23	$1/7$	$0.316 + 1/7 = 0.459$	$\exp(-0.459) = 0.632$
31	$1/5$	$0.459 + 1/5 = 0.659$	$\exp(-0.659) = 0.517$
34	$1/4$	$0.659 + 1/4 = 0.909$	$\exp(-0.909) = 0.403$
48	$1/2$	$0.909 + 1/2 = 1.409$	$\exp(-1.409) = 0.244$

7. *Using the R code provided calculate the Nelson-Aalen estimator and plot it and the Kaplan-Meier Curves. You can use the NA output to confirm your calculations in the previous problem.*

   The Kaplan-Meier and Nelson-Aalen curves are pretty similar, but the NA survival probabilities are a little bit bigger than the KM survival probabilities.

```
fitNA <- survfit(coxph(aml.obj[1:11] ~ 1, data = aml[1:11,]), type = 'aalen')
plot(fitNA$time, fitNA$surv, type = 's', ylim = c(0, 1),
     xlab = 'Survival Time', ylab = 'Suvival Probability')
points(fitNA$time, fitNA$surv, pch = 1)
lines(fitKM$time[1:9], fitKM$surv[1:9], type = 's', lty = 2)
points(fitKM$time[1:9], fitKM$surv[1:9], pch = 3)
legend(100, 0.8, c('Nelson-Aalen', 'Kaplan-Meier'), pch = c(1, 3), lty = c(1, 2))
```




```
summary(fitNA)

Call: survfit(formula = coxph(aml.obj[1:11] ~ 1, data = aml[1:11, ]),
    type = "aalen")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    9     11       1    0.913   0.083        0.764        1.000
   13     10       1    0.826   0.112        0.634        1.000
```

18	8	1	0.729	0.134	0.508	1.000
23	7	1	0.632	0.147	0.400	0.998
31	5	1	0.517	0.159	0.283	0.945
34	4	1	0.403	0.160	0.185	0.876
48	2	1	0.244	0.156	0.070	0.853

8. *Fit a Cox Proportional Hazards model to these data and summarize the results giving me an estimate of the appropriate hazard ratio and an associated 95% CI. Interpret the results in terms of the problem.*

```
fit.cox <- coxph(aml.obj ~ x, data = aml)
summary(fit.cox)

Call:
coxph(formula = aml.obj ~ x, data = aml)

  n= 23, number of events= 18

                coef exp(coef) se(coef)     z Pr(>|z|)
xNonmaintained 0.9155    2.4981   0.5119 1.788   0.0737

               exp(coef) exp(-coef) lower .95 upper .95
xNonmaintained     2.498     0.4003    0.9159     6.813

Concordance= 0.619  (se = 0.073 )
Rsquare= 0.137   (max possible= 0.976 )
Likelihood ratio test= 3.38  on 1 df,   p=0.06581
Wald test            = 3.2  on 1 df,   p=0.07371
Score (logrank) test = 3.42  on 1 df,   p=0.06454
```

We estimate the hazard of death for individuals whose chemotherapy is not maintained to be 2.4981 times the hazard of death for people whose chemotherapy is maintained. An approximate 95% confidence interval for the hazard ratio is 0.9159 to 6.813.