Homework 9

Qianbo Wang
uni: qw2180

Consider the colon data in the R package "survival". It gives adjuvant chemotherapy data for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals; 5-FU is a moderately toxic (as these things go) chemotherapy agent. There are two records per person, one for recurrence (etype=1) and one for death (etype=2). Other important variables include:

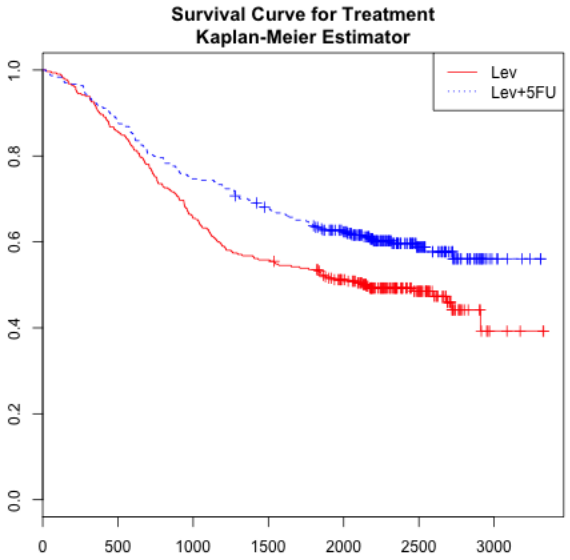
- **rx: Treatment: Obs(ervation), Lev(amisole), Lev(amisole)+5-FU**
- **sex: 1=male**
- **age: in years**
- **time: days until event or censoring**
- **status: censoring status**

For the following, consider survival to be Days until Death, i.e., etype=2.

Problem 1

Using the Kaplan-Meier method, estimate the survival curve for each treatment group: Lev(amisole) and Lev(amisole)+5-FU.

Using Kaplan-Meier method, fit the survival estimator, the estimated survival curve is as follows:



Problem 2

Estimate the median survival time for each of the two treatment groups, using the estimated survival curves.

Using Kaplan-Meier survival curve, the result is as follows:

Kaplan-Meier Estimated Curve

```
Call: survfit(formula = Surv(time, status) ~ rx, data = survival_colon)

             n events median 0.95LCL 0.95UCL
rx=Lev     310    161   2152    1540      NA
rx=Lev+5FU 304    123     NA    2725      NA
```

So the estimated median is:

Table 1: Estimated Median for two treatment groups

Treatment	Lev	Lev+5FU
Median	2152	NA

The median is calculated as the smallest survival time for which the survivor function is less than or equal to 0.5.

- For treatment Lev, the estimated median is 2152.
- For treatment Lev+5FU, since it never gets to the point where $s(t) \leq 0.5$, then the median for treatment Lev+5FU is NA.

Problem 3

Using the log-rank test, determine whether there is a difference in survival between the two groups.

Using Log-rank test, the null hypothesis and alternative hypothesis are:

H_0: there is no difference between the two treatment of survival curves

H_1: there is difference between the two treatment of survival curves

the result is as follows:

Log-rank Test Result

```
Call:
survdiff(formula = Surv(time, status) ~ rx, data = survival_colon)

             N Observed Expected (O-E)^2/E (O-E)^2/V
rx=Lev     310      161      137      4.24      8.21
rx=Lev+5FU 304      123      147      3.95      8.21

 Chisq= 8.2  on 1 degrees of freedom, p= 0.00417
```

Since the p-value of log-rank test is 0.00417, then we should reject the null hypothesis, i.e., concluding that there is a difference in survival between the two groups.

Problem 4

Using a Cox proportional hazards model, estimate the hazard ratio for Levamisole relative to 5-FU, adjusting for Age and Sex.

Using Cox proportional hazards model and adjusting for Age and Sex, the result is as follows:

Cox proportional hazards model adjusting for age and sex Result

```
Call:
coxph(formula = Surv(time, status) ~ rx_d + age + sex, data = survival_colon)

  n= 614, number of events= 284

          coef exp(coef)  se(coef)      z Pr(>|z|)
rx_d  0.344635  1.411474  0.120079  2.870   0.0041 **
age   0.000920  1.000921  0.005119  0.180   0.8574
sex  -0.043126  0.957791  0.118948 -0.363   0.7169
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

     exp(coef) exp(-coef) lower .95 upper .95
rx_d    1.4115     0.7085    1.1155     1.786
age     1.0009     0.9991    0.9909     1.011
sex     0.9578     1.0441    0.7586     1.209

Concordance= 0.549  (se = 0.018 )
Rsquare= 0.014   (max possible= 0.996 )
Likelihood ratio test= 8.38  on 3 df,   p=0.0388
Wald test            = 8.29  on 3 df,   p=0.04033
Score (logrank) test = 8.37  on 3 df,   p=0.03892
```

Since we want to get the hazard ratio for Levamisole relative to 5-FU, then we simply set the Lev+5FU as the baseline.

Since from the result we can see, the hazard ratio for the second group relative to the first group, that is, rxLev relative to rxLev+5FU with adjusting for age and sex is 1.4115.

Problem 5

Give a 95% confidence interval for the hazard ratio in 4.

From the result above, we can see that the 95% confidence interval for hazard ratio in Problem 4 is:

$$[1.1155, 1.786]$$

R Code:

```
rm(list = ls())
library(survival)
data("colon")

#Meier Estimator for treatment Lev, Lev+5FU
survival_colon <- subset(colon,etype==2 & rx %in% c("Lev","Lev+5FU"))
fit_trt<- survfit(formula = Surv(time, status) ~ rx, data = survival_colon, type = "kaplan-meier")
sink('/Users/raymond/Drive/STAT W4201/HW9/meiersum.txt')
summary(fit_trt)
sink()

#Plot the survival curve
png(filename = "/Users/raymond/Drive/STAT W4201/HW9/meiercurve.png")
plot(fit_trt, lty = c(1,2), col = c("red","blue"))
title(main="Survival Curve for Treatment")
title(main="Kaplan-Meier Estimator",line=0.5)
legend("topright",c("Lev","Lev+5FU"),col=c("red","blue"),lty=c(1,3))
dev.off()

#Estimate the median
sink('/Users/raymond/Drive/STAT W4201/HW9/meier.txt')
fit_trt
sink()

#Log-rank test for difference
log_rank<-survdiff(Surv(time, status) ~ rx, data=survival_colon)

sink('/Users/raymond/Drive/STAT W4201/HW9/logrank.txt')
log_rank
sink()

#Cox proportional hazards model
survival_colon$rx_d<-as.numeric(survival_colon$rx=="Lev")
cox<-coxph(Surv(time,status)~rx_d+age+sex,data=survival_colon)
sink('/Users/raymond/Drive/STAT W4201/HW9/cox.txt')
summary(cox)
sink()
```