

## Seet, Andrea- Epi Analysis HW #3

```
# Read in the data
library(knitr)
library(xtable)
library(foreign)
library(survival)
leuk <- read.dta("/Users/aseet/Documents/Fall 2019/Epi
Analysis/Data/leuk.dta")

survfit.leuk <- survfit(Surv(time,cens)~group,data=leuk)
summary(survfit.leuk)

## Call: survfit(formula = Surv(time, cens) ~ group, data = leuk)
##
##               group=6 MP
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    6      21       3   0.857  0.0764    0.720    1.000
##    7      17       1   0.807  0.0869    0.653    0.996
##   10      15       1   0.753  0.0963    0.586    0.968
##   13      12       1   0.690  0.1068    0.510    0.935
##   16      11       1   0.627  0.1141    0.439    0.896
##   22       7       1   0.538  0.1282    0.337    0.858
##   23       6       1   0.448  0.1346    0.249    0.807
##
##               group=Placebo
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1      21       2   0.9048  0.0641    0.78754    1.000
##    2      19       2   0.8095  0.0857    0.65785    0.996
##    3      17       1   0.7619  0.0929    0.59988    0.968
##    4      16       2   0.6667  0.1029    0.49268    0.902
##    5      14       2   0.5714  0.1080    0.39455    0.828
##    8      12       4   0.3810  0.1060    0.22085    0.657
##   11       8       2   0.2857  0.0986    0.14529    0.562
##   12       6       2   0.1905  0.0857    0.07887    0.460
##   15       4       1   0.1429  0.0764    0.05011    0.407
##   17       3       1   0.0952  0.0641    0.02549    0.356
##   22       2       1   0.0476  0.0465    0.00703    0.322
##   23       1       1   0.0000    NaN          NA          NA
```

Use Surv to set up your outcome object

```
s1_leuk <- with(leuk, Surv(time = time, event = cens))

summary_s1 <- summary(s1_leuk)
summary_s1
```

```
##      time      status
## Min.   : 1.00   Min.   :0.0000
## 1st Qu.: 6.00   1st Qu.:0.0000
## Median :10.50   Median :1.0000
## Mean    :12.88   Mean    :0.7143
## 3rd Qu.:18.50   3rd Qu.:1.0000
## Max.    :35.00   Max.    :1.0000
```

Produce a Kaplan-Meier table Kaplan-Meier estimate of Survival Function/Curve

```
s1_fit <- survfit(s1_leuk ~ 1, data = leuk, conf.type = "log-log")
s1_fit.summary <- summary(s1_fit)

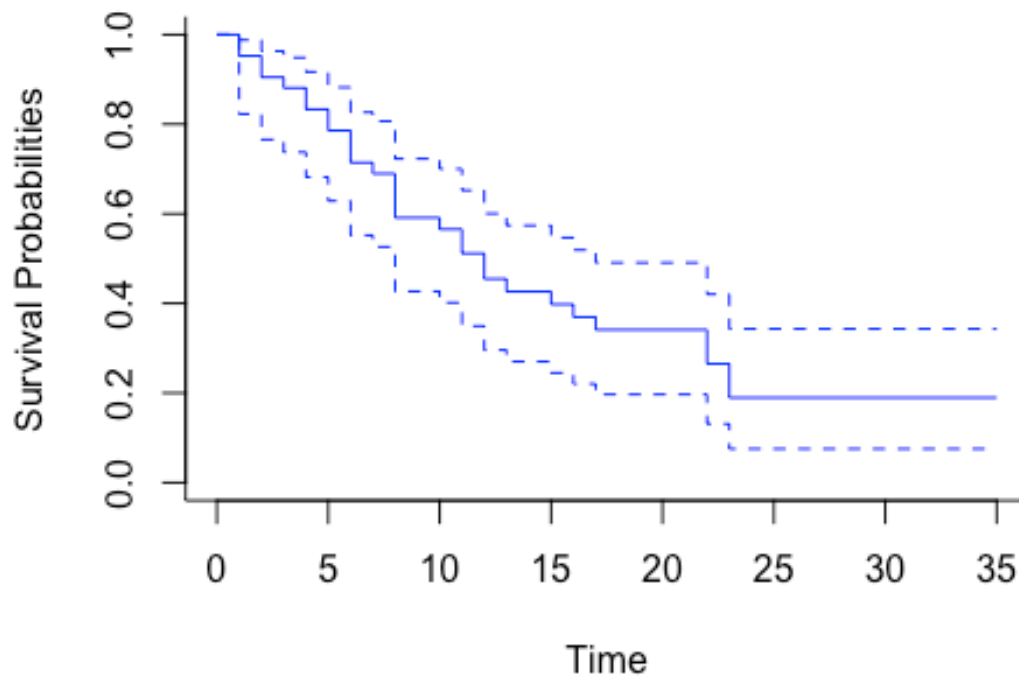
keep.cols <- c("time", "n.risk", "n.event", "n.censor", "surv", "std.err",
               "upper", "lower")
KMtable <- sapply(s1_fit.summary[keep.cols], cbind)
round(KMtable[1:17,], digits = 3)
```

	time	n.risk	n.event	n.censor	surv	std.err	upper	lower
## [1,]	1	42	2	0	0.952	0.033	0.988	0.823
## [2,]	2	40	2	0	0.905	0.045	0.963	0.766
## [3,]	3	38	1	0	0.881	0.050	0.949	0.737
## [4,]	4	37	2	0	0.833	0.058	0.917	0.682
## [5,]	5	35	2	0	0.786	0.063	0.882	0.629
## [6,]	6	33	3	1	0.714	0.070	0.826	0.552
## [7,]	7	29	1	0	0.690	0.072	0.807	0.526
## [8,]	8	28	4	0	0.591	0.076	0.723	0.427
## [9,]	10	23	1	2	0.565	0.077	0.700	0.402
## [10,]	11	21	2	1	0.512	0.079	0.652	0.349
## [11,]	12	18	2	0	0.455	0.080	0.601	0.296
## [12,]	13	16	1	0	0.426	0.080	0.574	0.270
## [13,]	15	15	1	0	0.398	0.079	0.547	0.245
## [14,]	16	14	1	0	0.369	0.078	0.519	0.220
## [15,]	17	13	1	1	0.341	0.077	0.491	0.197
## [16,]	22	9	2	2	0.265	0.077	0.420	0.131
## [17,]	23	7	2	0	0.189	0.071	0.343	0.075

Plot KM Survival Curve

```
plot(s1_fit,
     bty = "l",
     col = "blue",
     xlab = "Time",
     ylab = "Survival Probabilities",
     main = "Kaplan-Meier Survival Curve")
```

## Kaplan-Meier Survival Curve



Verify probabilities

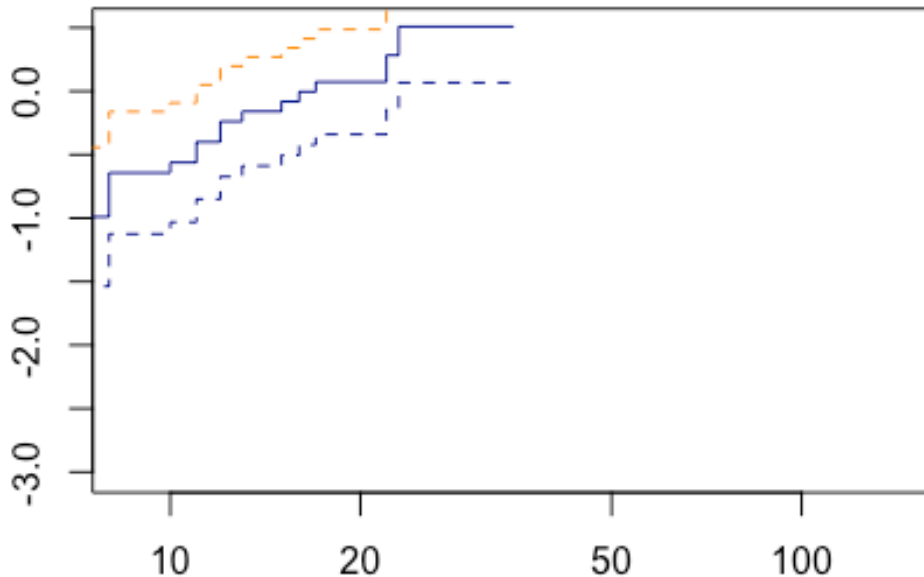
```
# LOGRANK TEST  
survdif(Surv(time, cens) ~ group, data=leuk)
```

Want to see non-crossing lines which we observe

```
plot(s1_fit,  
     fun = "cloglog",  
     col = c("darkblue", "darkorange"))
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 1 x value <= 0 omitted  
## from logarithmic plot
```

```
legend(x = 200, y = -4,  
       legend = c("untreated", "treated"),  
       col = c("darkblue", "darkorange"),  
       lty = 1)
```



6-MP  $S(10) = 0.807$   $S(20) = 0.627$   $S(30) = 0.448$

Placebo  $S(10) = 0.3810$   $S(20) = 0.0952$

Show that the estimated conditional probability of remission at week 20, conditional on being in remission at week 10, is approximately  $0.25 S(20)/S(10)$

$0.0952/0.3810$

```
## [1] 0.2498688
```

```
survfit.leuk <- survfit(Surv(time, cens) ~ group, data = leuk)
summary(survfit.leuk)
```

```
## Call: survfit(formula = Surv(time, cens) ~ group, data = leuk)
```

```
##
```

```
##          group=6 MP
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	6	21	3	0.857	0.0764	0.720	1.000
##	7	17	1	0.807	0.0869	0.653	0.996
##	10	15	1	0.753	0.0963	0.586	0.968
##	13	12	1	0.690	0.1068	0.510	0.935
##	16	11	1	0.627	0.1141	0.439	0.896
##	22	7	1	0.538	0.1282	0.337	0.858

```
##      23      6      1      0.448  0.1346      0.249      0.807
##
##                               group=Placebo
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     21      2  0.9048  0.0641    0.78754    1.000
##    2     19      2  0.8095  0.0857    0.65785    0.996
##    3     17      1  0.7619  0.0929    0.59988    0.968
##    4     16      2  0.6667  0.1029    0.49268    0.902
##    5     14      2  0.5714  0.1080    0.39455    0.828
##    8     12      4  0.3810  0.1060    0.22085    0.657
##   11      8      2  0.2857  0.0986    0.14529    0.562
##   12      6      2  0.1905  0.0857    0.07887    0.460
##   15      4      1  0.1429  0.0764    0.05011    0.407
##   17      3      1  0.0952  0.0641    0.02549    0.356
##   22      2      1  0.0476  0.0465    0.00703    0.322
##   23      1      1  0.0000    NaN          NA          NA
```

Estimate the probabilities of remaining in remission at 20 and 30 weeks for the 6-MP group  
 6-MP  $S(10) = 0.807$   $S(20) = 0.627$   $S(30) = 0.448$

Probability of remaining in remission at 20 weeks, condition on being in remission at 10 weeks is 77.7%  $S(20)/S(10)$

$0.627/0.807$

```
## [1] 0.7769517
```

Probability of remaining in remission at 30 weeks, condition on being in remission at 10 weeks is 55.5%  $S(30)/S(10)$

$0.448/0.807$

```
## [1] 0.5551425
```

$$\log[\lambda|albumin] = \log[h_o(t)] + \beta * albumin_1 - \log[\lambda|albumin]$$

$$= \log[h_o(t)] + \beta * albumin_0 = \beta * (albumin_1 - albumin_0)$$

replace albumin\_1 with each of the comparison levels Question of changing the reference level to 3

```
library(multcomp)
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: TH.data
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##      leuk

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##      geyser

pbc.data <- read.dta("/Users/aseet/Documents/Fall 2019/Epi
Analysis/Data/pbc.dta")
pbc.data$albumin[pbc.data$albumin == 0] <- 3
cox.pbc.1 <- coxph(Surv(years, as.integer(status=="Dead"))~albumin,
                  ties="efron", data=pbc.data)
summary(cox.pbc.1)

## Call:
## coxph(formula = Surv(years, as.integer(status == "Dead")) ~ albumin,
##       data = pbc.data, ties = "efron")
##
##      n= 312, number of events= 125
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## albumin -1.7957    0.1660   0.2094 -8.576  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## albumin    0.166     6.024   0.1101   0.2503
##
## Concordance= 0.711 (se = 0.025 )
## Likelihood ratio test= 63.95 on 1 df,  p=1e-15
## Wald test            = 73.54 on 1 df,  p=<2e-16
## Score (logrank) test = 70.89 on 1 df,  p=<2e-16
```

```
albumin = 2.5 exp(beta*(2.5-3.0))
```

```
exp(-1.8*-0.5)
```

```
## [1] 2.459603
```

```
albumin = 3.5
```

```
exp(-1.8*0.5)
```

```
## [1] 0.4065697
```

```
albumin = 4.0
```

```
exp(-1.8*1)
```

```
## [1] 0.1652989

#Create a
k <-matrix(c(2.5-3,3.5-3,4.0-3),nrow=3)
ln.hr <-coefficients(glht(cox.pbc.1, k))
exp(ln.hr)

##          1          2          3
## 2.4543200 0.4074448 0.1660113

cox.pbc.2<- coxph(Surv(years,as.integer(status=="Dead"))~cholest+bilirubin,
                  ties="efron",data=pbcc.data)
beta.pbc.2<-coefficients(cox.pbc.2)
exp(beta.pbc.2[1]*100)# HR per 100 mg/dL increase in cholesterol

##      cholest
## 0.9888254

exp(beta.pbc.2[2]*10)# HR per 10 mg/dL increase in bilirubin

## bilirubin
## 4.352349
```

For a 100mg/dL increase in cholesterol, the hazard of death decreases 2%. For a 10 mg/dL increase in bilirubin, the hazard of death increases by 4.4 times. or 440%.